frontiers | Research Topics

Current progress in genomic and genetic research on human viral diseases

Edited by
Binhua Liang, Birte Möhlendick, Hezhao Ji and Xiang Gao

Published in
Frontiers in Genetics
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-4833-2
DOI 10.3389/978-2-8325-4833-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Current progress in genomic and genetic research on human viral diseases

Topic editors

Binhua Liang — National Microbiology Laboratory, Public Health Agency of Canada (PHAC), Canada

Birte Möhlendick — Essen University Hospital, Germany

Hezhao Ji — National Microbiology Laboratory, Public Health Agency of Canada (PHAC), Canada

Xiang Gao — Loyola University Chicago, United States

Citation

Liang, B., Möhlendick, B., Ji, H., Gao, X., eds. (2024). *Current progress in genomic and genetic research on human viral diseases*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4833-2

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 15 April 2024
DOI 10.3389/fgene.2024.1407559

OPEN ACCESS

EDITED AND REVIEWED BY
Maxim B. Freidin,
Queen Mary University of London,
United Kingdom

*CORRESPONDENCE
Hezhao Ji,
hezhao.ji@phac-aspc.gc.ca
Binhua Liang,
ben.liang@phac-aspc.gc.ca

RECEIVED 26 March 2024
ACCEPTED 01 April 2024
PUBLISHED 15 April 2024

CITATION
Ji H, Möhlendick B, Gao X and Liang B (2024), Editorial: Current progress in genomic and genetic research on human viral diseases. *Front. Genet.* 15:1407559. doi: 10.3389/fgene.2024.1407559

Editorial: Current progress in genomic and genetic research on human viral diseases

Hezhao Ji[1,2]*, Birte Möhlendick[3], Xiang Gao[4] and Binhua Liang[5,6]*

[1]National Microbiology Laboratory at JC Wilt Infectious Diseases Research Centre, Public Health Agency of Canada, Winnipeg, MB, Canada, [2]Department of Medical Microbiology and Infectious Diseases, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB, Canada, [3]Institute of Pharmacogenetics, Essen University Hospital, Essen, Germany, [4]Department of Medicine, Stritch School of Medicine, Loyola University Chicago, Maywood, IL, United States, [5]Medical and Scientific Affairs, National Microbiology Laboratory Branch, Public Health Agency of Canada, Winnipeg, MB, Canada, [6]Department of Biochemistry and Medical Genetics, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB, Canada

KEYWORDS

genomics, genetics, human viral diseases, immunogenetics, viral evolution

Editorial on the Research Topic
Current progress in genomic and genetic research on human viral diseases

Viral diseases pose a global public health threat, demanding a deeper understanding of the intricate interplay between viruses and their human hosts. The recent COVID-19 pandemic underscores the persistent challenges in combating viral pathogens and highlights the limited comprehension of their complex interactions. Insightful genomic and genetic studies are crucial in unraveling the co-evolution of viruses and hosts during infections. Such studies either delve into human genetics to investigate determinants of immune responses against viruses or explore viral genetic diversity that contributes to viral fitness, virulence, immune evasion, and viral evolution. The integration of both host and virus perspectives enhances our comprehension of the human-virus interplay, paving the way for improved preventative and therapeutic strategies. Recent advances in next-generation sequencing, bioinformatics, and data science have significantly propelled genetic and genomic studies, marking a promising era in our pursuit of combating viral infections.

This Research Topic invited a wide range of contributions, including original research, reviews, and perspectives, focusing on genetics and genomics in the context of human viral diseases. By emphasizing different aspects, the Research Topic encouraged submissions exploring evolutionary genetics to address viral evolution, genome analysis, and comparative genomics of human viruses. Key areas of interest encompassed genetic determinants influencing human susceptibility to viral infections, the impact of host genetics on vaccination response, and the exploration of viral genetics concerning infectivity, tropism, immune evasion, and antiviral resistance. Additionally, the Research Topic also welcomed new methodologies addressing viral genetic diversity and its role in viral pathogenesis. By promoting a comprehensive understanding of these genetic and genomic aspects, this Research Topic aimed to advance our knowledge and contribute to the development of effective strategies against human viral diseases.

We are delighted to have received many excellent submissions, nine of which have been successfully published in this Research Topic. This editorial piece may serve as a brief overview of these articles, highlighting their substantial contributions to the Research Topic.

This compilation includes three original research articles addressing the pressing Research Topic of COVID-19 and SARS-CoV-2. Čiučiulkaitė, et al. focused on the effect of the human *GNB3* c.825C>T polymorphism on immune responses upon COVID-19 mRNA vaccination. This was one of the few studies at the time to document the significant decline in antibody titers and T-cell responses from months 1–6 after the second dose of vaccination. It also was one of the few studies to examine the kinetics of T-cell responses against SARS-CoV-2 after mRNA vaccination. While the *GNB3* c.825C>T polymorphism had no significant effect on the humoral immune response, individuals with the CC genotype at this locus exhibited significantly enhanced T-cell responses after mRNA-1273 vaccination, suggesting improved protection against COVID-19. Möhlendick, et al. from the same research group explored the potential impact of the *GNB3* c.825C>T polymorphism on the clinical outcome of COVID-19. This study, which included 1,570 individuals, examined the potential associations between demographic factors, pre-existing conditions, laboratory parameters, *GNB3* rs5443 genotype, and COVID-19 clinical outcome. Notably, in addition to identifying associating factors against fatal COVID-19 outcomes among other examined parameters, they reported that the *GNB3* rs5443 TT genotype was significantly associated with a 40% reduced risk of COVID-19 fatality, suggesting its potential as a prognostic biomarker. While the findings of Möhlendick et al. on *GNB3* genotypes may seem to contradict the observations of Čiučiulkaitė, et al., it is important to note that the former focused on clinical outcomes in unvaccinated individuals, while the latter focuses on immune responses to COVID-19 vaccination. These divergent findings underscore the complexity of the interaction between SARS-CoV-2 and the host genetics and immune responses, and further research is needed to reconcile and fully understand these relationships.

Another COVID-19-related study was conducted by Li et al., who investigated into the SARS-CoV-2 pathogenesis from a viral genetics perspective. Among many SARS-CoV-2 variants of concern (VOC), the Delta variant emerged from India in 2020 and swept the world in mid-2021 with a high transmission rate and high mortality. This study investigated the evolution of the Delta variant by analyzing mutational patterns in viral genomes originating from India at three different time periods surrounding the emergence of Delta VOC in the country. The study revealed a progressive increase in viral mutations, with viral genomes exhibiting the highest diversity following the establishment of Delta VOC. Notably, the pre-Delta phase showed a higher number of negatively selected sites, which may have protected critical gene regions during evolution. This study suggests ongoing viral adaptation and evolution and highlights the dynamic genetic landscape of SARS-CoV-2.

Not surprisingly, HIV-1 is another virus heavily studied in this Research Topic of publications. Yang et al. identified and characterized two novel unique recombinant forms (URFs) of HIV-1 from Hebei, China. Based on the analysis of near full-length genome (NFLG) sequences, both URFs were shown to result from the recombination of CRF01_AE and CRF07_BC, but with distinct recombination breakpoints. Similarly, Zhang et al. reported two novel HIV-1 URFs resulting from HIV-1 CRF01_AE and subtype B recombination in two MSM patients. Xing et al. focused their study on the origin and spread of another novel CRF - CRF68_01B initially discovered among men who have sex with men (MSM) populations. Phylodynamic analysis of CRF68_01B helped to trace the origin of CRF68_01B back to Shenzhen in 2003, indicating subsequent spread to other regions. Molecular network analysis further revealed interprovincial transmission and highlighted the significant role of MSM populations in the spread of CRF68_01B. Fan et al. conducted a study investigating integrase strand transfer inhibitor (INSTI) resistance mutations in a treatment-naive HIV-positive patient cohort in Hebei, China. This study, pertinent to the global use of INSTI-containing antiretroviral therapy, identified both major and accessory INSTI-resistance mutations, with an overall INSTI resistance rate of 3.82%. Although the frequency of INSTI resistance was low, this study highlights the need for pre-treatment testing and increased resistance monitoring during INSTI-based ART regimens. Overall, these studies have provided valuable insights into the genomic diversity, transmission dynamics, and drug resistance profiles of HIV-1, highlighting the importance of ongoing research in understanding and managing this complex viral infection.

This Research Topic includes two review articles addressing the ongoing evolution of SARS-CoV-2 and the application of quantitative metagenomics next-generation sequencing (Q-mNGS) in the detection of infectious agents causing fever of unknown origin (FUO). The review by Fang et al. outlines the trajectory evolution of SARS-CoV-2, focusing on the amino acid variations of the spike protein and genomic recombination. Pre-Omicron variants showed concentrated spike protein mutations, including early D614G, which alters the antigenicity, transmissibility, and pathogenicity of SARS-CoV-2, while Omicron introduced numerous novel mutations, enhancing transmissibility and immune evasion without increasing clinical severity. The recombinant XBB variant, emerged in 2022, likely resulted from co-circulation and co-infection in immunocompromised patients. Despite transmission advantages, these variants have demonstrated moderate antibody escape, necessitating increased surveillance for genomic variation, particularly spike protein mutations and recombination, broad-spectrum therapeutics, and widespread vaccination efforts. Dong et al. reviewed the application of Q-mNGS as a transformative method for the detection of infectious agents causing FUO. This high-throughput sequencing technology outperforms conventional molecular diagnostic methods, providing faster and more comprehensive results at a lower cost, potentially revolutionizing FUO evaluation and reducing unnecessary testing.

Collectively, we are convinced that the compilation of publications within this Research Topic provides a glimpse into the current focus and research progress in genomic and genetic studies of human viral diseases.

Author contributions

HJ: Writing–original draft. BM: Writing–review and editing. XG: Writing–review and editing. BL: Writing–review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors would like to thank all the contributing authors, peer reviewers, and the editorial staff for their great contributions and generous support in the development and implementation of this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Microbiology

REVIEW
published: 04 July 2022
doi: 10.3389/fmicb.2022.931058

Use of Quantitative Metagenomics Next-Generation Sequencing to Confirm Fever of Unknown Origin and Infectious Disease

*Yuxin Dong†, Yulei Gao†, Yanfen Chai and Songtao Shou**

Department of Emergency Medicine, Tianjin Medical University General Hospital, Tianjin, China

A body temperature >38.3°C that lasts ≥3 weeks and lacks a clear diagnosis after 1 week of standard hospital examination and treatment is called "fever of unknown origin" (FUO). The main causes of FUO are infections, hematological diseases, autoimmune diseases, and other non-infectious inflammatory diseases. In recent years, quantitative metagenomics next-generation sequencing (Q-mNGS) has been used widely to detect pathogenic microorganisms, especially in the contribution of rare or new (e.g., severe acute respiratory syndrome-coronavirus-2) pathogens. This review addresses the undetermined cause of fever and its evaluation by Q-mNGS.

Keywords: quantitative metagenomics next-generation sequencing, fever of unknown origin, infections, pathogen, rare disease

OPEN ACCESS

Edited by:
Hezhao Ji,
Public Health Agency of Canada (PHAC), Canada

Reviewed by:
Eswarappa Pradeep Bulagonda, Sri Sathya Sai Institute of Higher Learning (SSSIHL), India
Binhua Liang,
Public Health Agency of Canada (PHAC), Canada

***Correspondence:**
Songtao Shou
stshou@qq.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *28 April 2022*
Accepted: *15 June 2022*
Published: *04 July 2022*

Citation:
Dong Y-x, Gao Y-l, Chai Y-f and Shou S-t (2022) Use of Quantitative Metagenomics Next-Generation Sequencing to Confirm Fever of Unknown Origin and Infectious Disease. Front. Microbiol. 13:931058. doi: 10.3389/fmicb.2022.931058

INTRODUCTION

Fever of unknown origin (FUO) can be caused by various diseases. More than 200 causes of FUO have been reported (Horowitz, 2013). In 1961, Petersdorf and Beeson were the first to define FUO as a state of febrile illness for more than 3 weeks, with a body temperature greater than 38.3°C (101°F) on several occasions and an uncertain diagnosis after 1 week of standard hospital examination and treatment (Petersdorf and Beeson, 1961). In 1991, Durak and Street re-defined FUO into four groups: "classic," "nosocomial," "neutropenic," and "human immunodeficiency virus (HIV)-associated." They proposed three outpatient visits and related investigations as an alternative to "1 week of hospitalization" (Durack and Street, 1991). In 1997, Arnow and Flaherty updated the FUO definition and considered the "minimum diagnostic evaluation to qualify as FUO" to be: comprehensive history-taking; repeated physical examination; complete blood count (including differential and platelet counts); routine blood chemistry (including lactate dehydrogenase, bilirubin, and liver enzymes); urinalysis (including microscopic examination); chest radiograph; erythrocyte sedimentation rate (ESR); antinuclear antibodies; rheumatoid factor; angiotensin-converting enzyme; routine blood cultures (×3) while not receiving antibiotics; cytomegalovirus immunoglobulin-M antibodies or virus detection in blood; heterophile antibody test in children and young adults; tuberculin skin test; computed tomography (CT) of the abdomen or radionuclide scan; HIV antibodies or virus-detection assay; further evaluation of abnormalities detected by the tests stated above (Arnow and Flaherty, 1997). Because of the complicated clinical characteristics and lack of laboratory indicators of a disease, the diagnosis is difficult and contributes to a high cost of hospitalization.

Infections, neoplasms, non-infectious inflammatory diseases, and other conditions are the four primary etiological groups for FUO (Mourad et al., 2003). Obtaining a detailed medical history and undertaking examinations to evaluate the cause of fever are crucial. The standard diagnosis and treatment process of FUO have not yet been proposed, but they should be performed in a specific order when carrying out the examination and diagnosis (**Figure 1**). Identification of pathogenic bacteria is crucial for targeted anti-infective medication (Haidar and Singh, 2022; Messacar et al., 2017). In patients with prolonged fever, empiric therapy is not recommended because it can mask symptoms, delay the diagnosis, and obstruct decision-making regarding optimal treatment (Unger et al., 2016). Only a few exceptions exist if treatment must be initiated based solely on diagnostic suspicion: antibiotics for culture-negative endocarditis, tuberculostatic agents for active tuberculosis, and glucocorticoids for temporal arteritis with a risk of vision loss (Bryan and Ahuja, 2007). Culture and testing of body fluids are common for a microbial diagnosis, but such cultures and tests are positive in only ∼40% of cases. Also, implementation and interpretation of blood cultures require time, which delays the information obtained by clinicians (Tromp et al., 2012). The sensitivity and specificity of PCR-based detection is based on the genomic sequence of known pathogenic bacteria, provides limited information, and is suboptimal for detection of mixed infections (Reuwer et al., 2019). Medication mistakes or treatment delays may arise due to the limits of clinical testing.

According to two systematic reviews conducted from 1995 to 2004 (Gaeta et al., 2006) and 2005 to 2015 (Fusco et al., 2019), infections are the leading cause of FUO. Screening and diagnostic processes must be developed to detect the pathogens that cause infection-related FUO. Quantitative metagenomics next-generation sequencing (Q-mNGS) is a current method to detect infection-related FUO pathogens. Quantitative metagenomics next-generation sequencing, also known as "high-throughput sequencing" or "massive parallel sequencing," is a type of technology that allows for the simultaneous and independent sequencing of hundreds to billions of DNA fragments (Morganti et al., 2019). Q-mNGS has many uses in clinical microbiological testing, and provides an unbiased method for pathogen detection. Recent studies have shown that Q-mNGS could be used to diagnose various infectious diseases, including coronavirus disease 2019 (COVID-19) (Ren et al., 2020), pneumonia due to *Chlamydia psittaci* infection (Chen et al., 2020), Ebola virus (EBOV) infection (Li et al., 2019), and talaromycosis (Shi et al., 2021).

Revolution in DNA-Sequencing: From Sanger Sequencing to Quantitative Metagenomics Next-Generation Sequencing

The "first generation" of gene-sequencing technology was born with the advent of the chain-termination method described by Sanger and Coulson (1975) and the chain-degradation method described by Maxam and Gilbert (1977). Gilbert and Sanger built the first sequencer in 1977 and used it to sequence the first

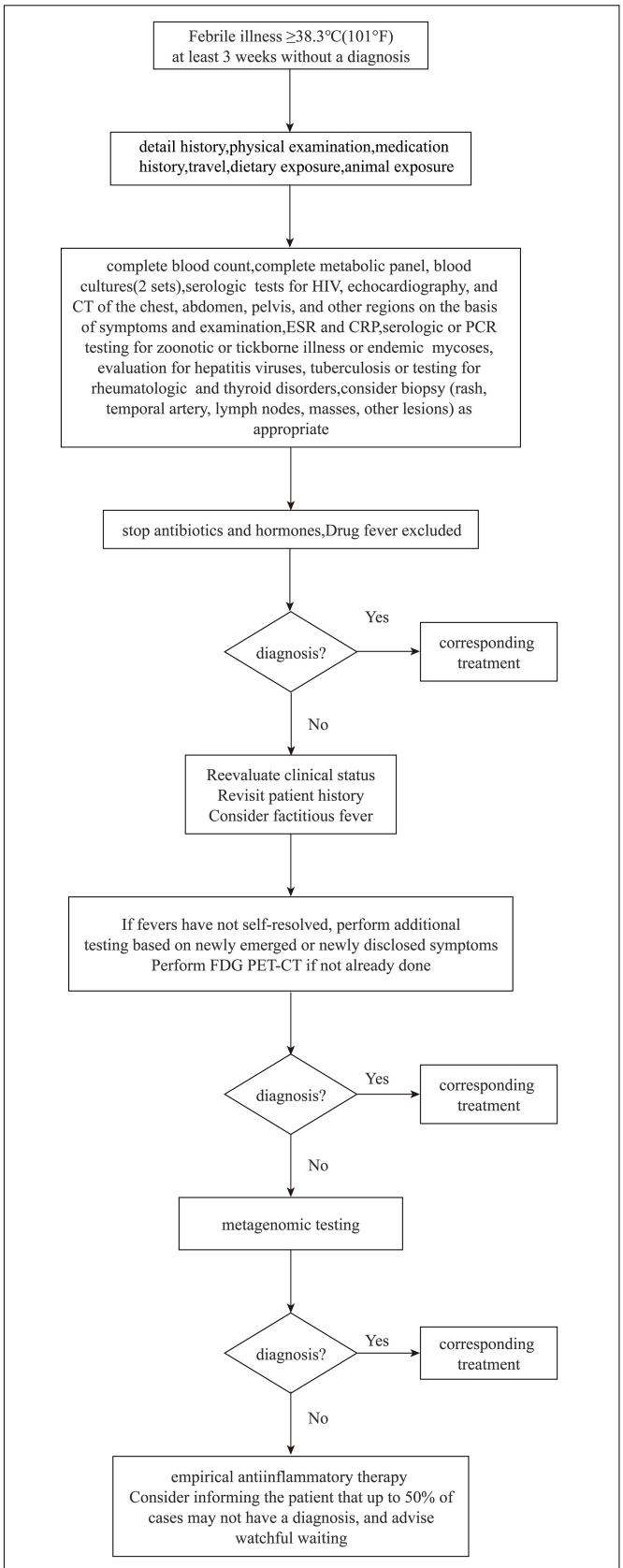


FIGURE 1 | FUO diagnosis and treatment flow chart.

full-length genome, phage X174, with 5,375 bases (Maxam and Gilbert, 1977; Aoyama et al., 1981). First-generation sequencing can produce a sequence of 700–1,000 bases at a time, so it cannot keep up with the pressing need for biological gene sequences.

Following a revolution in traditional sequencing technology, second-generation sequencing technology, known as "next-generation sequencing" (NGS), can be employed to obtain the sequences of hundreds of thousands to millions of nucleic-acid molecules in a single run. With NGS introduction, the transcriptome and genome of a species can be investigated in great detail. Jonathan Rothberg developed the biotechnology company 454 Life Sciences (Branford, CT, United States) in 2005 (Margulies et al., 2005). Other technologies, such as sequencing by oligonucleotide ligation and detection (SOLiD) (Applied Biosystems, Foster City, CA, United States) and Solexa (Illumina, San Diego, CA, United States), emerged subsequently. A total of 454 Life Sciences was acquired by Roche (Basel, Switzerland) in 2007.

The basic principles of this technology are that the DNA fragment does not need to be fluorescently labeled, there is no need for electrophoresis, and the sequence is changed by synthesis. A pyrophosphate group is removed when a base is added to the sequence, so pyrosequencing is also known as the detection of pyrophosphate bases (Nyrén et al., 1993; Ronaghi et al., 1998). Sequencing by SOLiD technology is based on ligase sequencing (Pérez-Enciso and Ferretti, 2010). Solexa technology (which is also used for sequencing-by-synthesis) was developed first by Illumina and is now used by the second-generation sequencer developed by Illumina (Pérez-Enciso and Ferretti, 2010; Strub et al., 2011).

Metagenomics (also known as "microbial environmental genomics") creates a metagenomic "library" by extracting the DNA or RNA of all microorganisms from environmental samples directly and studying them using genomics research strategies. Metagenomics based on NGS has become the focus of clinical research since the development of gene-sequencing technology.

Q-mNGS is a method for analyzing the genetic material of microbes and hosts from patient samples to diagnose infectious diseases. Q-mNGS has become the focus of clinical research due to the rapid advancement of gene-sequencing technology.

Third-generation sequencing technology includes the Pacific Bioscience (Levene et al., 2003) and Oxford Nanopore (Eisenstein, 2012) platforms, which are single-molecule technologies. Single-molecule sequencing (which does not require PCR amplification and which can, theoretically, determine nucleic-acid sequences of any length) is most notable when compared with first-generation and second-generation sequencing technologies (**Figure 2**).

Quantitative Metagenomics Next-Generation Sequencing in Fever of Unknown Origin or Infectious Diseases

The diagnostic value of NGS has been investigated in retrospective studies for patients suffering from fever (**Table 1**). The effectiveness of detection of NGS is higher than that of traditional methods. Fu et al. (2021) undertook a retrospective study on 175 patients with FUO to compare Q-mNGS with culture and traditional methods, including smears, serological tests, and amplification of nucleic acids (traditional PCR, Xpert MTB/RIF, and Xpert MTB/RIF Ultra). In comparison with culture and conventional methods, the authors concluded that Q-mNGS of blood might increase the overall rate of detection of novel organisms by 22.9 and 19.79%, and enhance the diagnostic rate by 38.0 and 32.0%, respectively. Zou et al. (2022) evaluated 12 patients with tuberculosis following renal transplantation, and Q-mNGS was helpful in 67% of cases.

Benamu et al. (2021) evaluated 55 patients with febrile neutropenia to compare the results of blood culture and standard microbiological testing within 24 h of fever onset and every 2–3 days. The Karius microbial cell-free DNA sequencing test (KT) sensitivity and specificity were 85% (41/48) and 100% (14/14), respectively. The calculated time-to-the-diagnosis was,

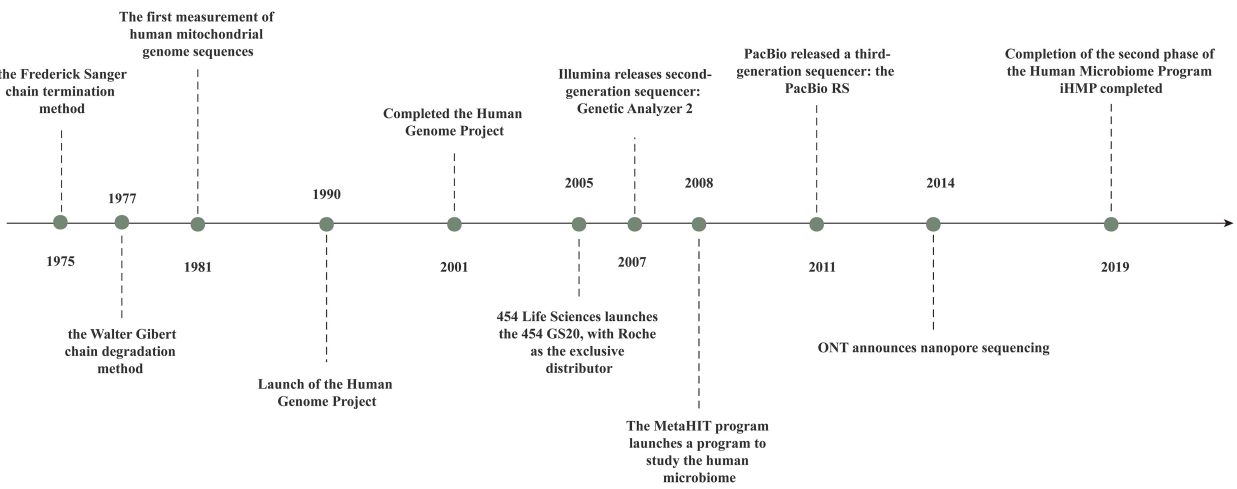


FIGURE 2 | The history of gene sequencingtechnology.

TABLE 1 | List of sequencing associated with fever or infectious disease' study.

References	Research type	Disease type	Method	Conclusion
Li et al. (2019)	Retrospective	70 patients of suspected Ebola hemorrhagic fever	mNGS vs. qRT-PCR_{US}	These results demonstrate the utility of mNGS in broad-based pathogen detection and outbreak surveillance.
Fu et al. (2021)	Retrospective	175 patients of FUO	mNGS vs. culture or traditional methods [smears, serological tests, nucleic acid amplification testing (NAAT)]	mNGS had significantly higher diagnostic efficacy in the FUO than culture or other traditional methods.
Zou et al. (2022)	Retrospective	12 patients of mycobacterium tuberculosis infection	Interferon-gamma release assay and NGS vs. the traditional PPD test and M. tuberculosis detection	The interferon-gamma release assay and NGS are relatively new detection methods with high sensitivity and specificity and can help with early TB diagnosis.
Benamu et al. (2021)	Prospective	55 patients with febrile neutropenia (FN)	The Karius microbial cell-free DNA (mcfDNA) sequencing Test (KT) vs. blood culture (BC) and standard microbiological testing (SMT)	The use of KT in the diagnosis and treatment of FN shows promise.
Liu et al. (2020)	Retrospective	17 patients of underwent lung transplantation	NGS vs. the bacteria culture method	NGS showed more sensitivity than bacterial culture for the detection of bacteria.
Reyes et al. (2021)	Retrospective	38 patients of febrile illness	mNGS vs. conventional viral pathogen detection methods (such as PCR)	In international travelers with febrile syndrome, viral metagenomics has the potential to help identify viral pathogens and co-infections in a single step.
Xiao et al. (2020)	Retrospective	8 patients of COVID-19	Meta sequencing vs. multiplex PCR amplification (amplicon) and hybrid capture (capture)	Meta-sequencing can be prioritized if other genetic materials are to be studied, such as target viruses that have become highly diversified through recombinational events, or if the viral load within the RNA sample is high.
Jerome et al. (2019)	Retrospective	40 patients of fever after traveling	mNGS analysis vs. standard of care diagnostics	MNGS has the potential to improve infectious disease diagnostic yield and detect multiple pathogens in a single sample.
Williams et al. (2018)	Retrospective	12 plasma specimens from patients with unexplained febrile illness	Unbiased sequencing vs. VirCapSeq-VERT (a positive selection system).	The utility of high-throughput sequencing strategies in outbreak investigations
Horiba et al. (2021)	Retrospective	112 patients of pediatric febrile neutropenia	NGS vs. blood cultures	NGS technique has great potential for detecting causative pathogens in patients with FN and may be effective for detecting pathogens in minute quantities of microbiota.

in general, shorter with KT (87%). Adjudicators determined real-time KT results have allowed early optimization of antimicrobial agents in 47% of patients. Liu et al. (2020) retrospectively evaluated 17 patients who received a lung transplant. The proportion of bacteria detected in the lungs of donors was 52.9 and 35.3% by NGS and bacterial culture, respectively. NGS was more sensitive for bacterial detection than the classic bacterial culture. Reyes and colleagues Xiao undertook a retrospective study on 38 patients. In eight of the 38 patients (21%), all viral pathogens detected by 42 conventional assays were also detected by Q-mNGS, and Q-mNGS resulted in additional pathogenic findings in two patients (5%).

NGS provides more information than conventional diagnostic tests. Xiao et al. (2020) were the first to systematically investigate inter- and intra-individual variations in severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) using amplicon- and capture-based whole-genome sequencing; it was also the first comparative study using multiple approaches. The study illustrated that ultra-high-throughput metatranscriptomic (meta) sequencing uncovered rich genetic information in clinical samples besides SARS-CoV-2, and provided references for clinical diagnostics and therapeutics. In June 2020, the US Food and Drug Administration granted Emergency Use Authorization for a Q-mNGS test for COVID-19 manufactured by Illumina, the first such authorization for a NGS diagnostic News in Brief (2020). Li et al. (2019) demonstrated that Q-mNGS of field-collected samples could be used to recover nine genomes from the EBOV outbreak in Boende (Democratic Republic of Congo) in 2014 (>50% coverage), detect the EBOV with a high sequencing depth of 17.3 ± 4.7 *SD* million reads with comparable sensitivity to PCR, and identify co-infections from well-recognized (*Plasmodium falciparum*) and novel/uncommon (e.g., Orungo virus) pathogens. Jerome et al. (2019) prospectively included 40 returning travelers presenting with fever (≥38°C) whose plasma samples were sequenced: 11 of 40 patients were diagnosed with a viral infection. Five viral infections were

detected by Q-mNGS that were also revealed in standard-of-care diagnostics, but two patients infected with the Chikungunya virus and one patient with the mumps virus were also diagnosed by Q-mNGS only. Williams et al. (2018) investigated the plasma virome from cases of unexplained febrile illness in Tanzania from 2013 to 2014 by sequencing methods. The latter could aid detection of viral coinfections, such as the nearly complete genomes of dengue virus-2 and human pestivirus. Horiba et al. (2021) evaluated 87 patients with febrile neutropenia. Putative pathogens were detected by Q-mNGS in 17.2% of patients, but all had negative blood cultures. Pathogenic detection methods (e.g., PCR) require clinicians to first suspect a specific bacterial infection before carrying out the corresponding detection. However, NGS technology can be employed to detect pathogenic bacteria in patient samples with high sensitivity, thereby providing recommendations for clinical treatment.

Often, NGS has been undertaken without using a structured diagnostic protocol, and at different stages of FUO. Zhu et al. (2021) found that use of Q-mNGS for blood as the first-line investigation could increase the diagnosis rate of FUO by 10.9% compared with that using culture, and that using Q-mNGS as the second-line investigation could improve the diagnosis rate of concurrent infection by 66.7 and 12.5% for non-bloodstream infection.

Ultimately, the cost of FUO assessment can be reduced by Q-mNGS application because the diagnosis will be achieved early because unnecessary and costly diagnostic tests will not be carried out. Chai et al. (2018) investigated the cost–benefit relationship of Q-mNGS in FUO in which a cause could not be found despite appropriate investigations. A decision tree was created to describe systematically the costs and benefits associated with NGS introduction. Each diagnostic pathway was made until a first- or second-line investigation was positive. NGS was introduced into the pathway as a supplement to first- or second-line investigations. Chai and colleagues reported NGS use as the first-line investigation assuming a probability of detecting the cause of cost-effectiveness in all cases of $\geq$60% using unit costs of diagnostic tests and procedures in Singapore dollars in 2016. In that analysis, using a rational set of rates for a second-line investigation, the total expected cost of using NGS as a second-line investigation was greater than that using it as a first-line investigation. Although that analysis excluded the costs associated with hospitalization duration, the faster and more definitive answers provided by NGS may enable additional cost savings.

CONCLUSION

Q-mNGS is a sensitive diagnostic method for FUO evaluation. It could become a routine procedure in the diagnostic workup of FUO. Q-mNGS appears to be cost-effective in FUO because it avoids unnecessary investigations and reduces the duration of hospitalization.

AUTHOR CONTRIBUTIONS

Y-LG and S-TS conceived this study. Y-XD and Y-LG collected clinical data and were responsible for patient care, and drafted the manuscript. Y-LG, S-TS, and Y-FC revised the manuscript critically. All authors have revised the final version of the manuscript and approved it for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China (81871593 and 81701931). This work was also funded by the Tianjin Key Medical Discipline (Specialty) Construction Project.

ACKNOWLEDGMENTS

We thank Shuzhang Cui (Emergency Department of Tianjin Medical University General Hospital) for help in study design.

REFERENCES

Aoyama, A., Hamatake, R. K., and Hayashi, M. (1981). Morphogenesis of Phi X174: in vitro synthesis of infectious phage from purified viral components. *Proc. Natl. Acad. Sci. U.S.A.* 78, 7285–7289.

Arnow, P. M., and Flaherty, J. P. (1997). Fever of unknown origin. *Lancet* 350, 575–580.

Benamu, E., Gajurel, K., Anderson, J. N., Lieb, T., Gomez, C. A., Seng, H., et al. (2021). Plasma microbial cell-free DNA next generation sequencing in the diagnosis and management of febrile neutropenia. *Clin. Infect. Dis.* 74, 1659–1668. doi: 10.1093/cid/ciab324

Bryan, C. S., and Ahuja, D. (2007). Fever of unknown origin: is there a role for empiric therapy? *Infect. Dis. Clin. North Am.* 21, 1213–1220. doi: 10.1016/j.idc.2007.08.007

Chai, J. H., Lee, C. K., Lee, H. K., Wong, N., Teo, K., Tan, C. S., et al. (2018). Cost-benefit analysis of introducing next-generation sequencing (metagenomic) pathogen testing in the setting of pyrexia of unknown origin. *PLoS One* 13:e0194648. doi: 10.1371/journal.pone.0194648

Chen, X., Cao, K., Wei, Y., Qian, Y., Liang, J., Dong, D., et al. (2020). Metagenomic next-generation sequencing in the diagnosis of severe pneumonias caused by *Chlamydia psittaci*. *Infection* 48, 535–542. doi: 10.1007/s15010-020-01429-0

Durack, D. T., and Street, A. C. (1991). Fever of unknown origin – reexamined and redefined. *Curr. Clin. Top. Infect. Dis.* 11, 35–51.

Eisenstein, M. (2012). Oxford Nanopore announcement sets sequencing sector abuzz. *Nat. Biotechnol.* 30, 295–296. doi: 10.1038/nbt0412-295

Fu, Z.-F., Zhang, H.-C., Zhang, Y., Cui, P., Zhou, Y., Wang, H.-Y., et al. (2021). Evaluations of clinical utilization of metagenomic next-generation sequencing in adults with fever of unknown origin. *Front. Cell. Infect. Microbiol.* 11:745156. doi: 10.3389/fcimb.2021.745156

Fusco, F. M., Pisapia, R., Nardiello, S., Cicala, S. D., Gaeta, G. B., and Brancaccio, G. (2019). Fever of unknown origin (FUO): which are the factors influencing the final diagnosis? A 2005–2015 systematic review. *BMC Infect. Dis.* 19:653. doi: 10.1186/s12879-019-4285-8

Gaeta, G. B., Fusco, F. M., and Nardiello, S. (2006). Fever of unknown origin: a systematic review of the literature for 1995-2004. *Nucl. Med. Commun.* 27, 205–211. doi: 10.1097/00006231-200603000-00002

Haidar, G., and Singh, N. (2022). Fever of unknown origin. *N. Engl. J. Med.* 386, 463–477. doi: 10.1056/NEJMra2111003

Horiba, K., Torii, Y., Okumura, T., Takeuchi, S., Suzuki, T., Kawada, J.-I., et al. (2021). Next-generation sequencing to detect pathogens in pediatric febrile neutropenia: a single-center retrospective study of 112 cases. *Open Forum Infect. Dis.* 8:ofab223. doi: 10.1093/ofid/ofab223

Horowitz, H. W. (2013). Fever of unknown origin or fever of too many origins? *N. Engl. J. Med.* 368, 197–199. doi: 10.1056/NEJMp1212725

Jerome, H., Taylor, C., Sreenu, V. B., Klymenko, T., Filipe, A. D. S., Jackson, C., et al. (2019). Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers. *J. Infect.* 79, 383–388. doi: 10.1016/j.jinf.2019.08.003

Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., and Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* 299, 682–686.

Li, T., Mbala-Kingebeni, P., Naccache, S. N., Thézé, J., Bouquet, J., Federman, S., et al. (2019). Metagenomic next-generation sequencing of the 2014 Ebola virus disease outbreak in the democratic republic of the Congo. *J. Clin. Microbiol.* 57:e00827-19. doi: 10.1128/JCM.00827-19

Liu, D., Zhang, J., Wu, B., Liu, F., Ye, S., Wang, H., et al. (2020). Impact of donor lung colonized bacteria detected by next-generation sequencing on early post-transplant outcomes in lung transplant recipients. *BMC Infect. Dis.* 20:689. doi: 10.1186/s12879-020-05393-w

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380. doi: 10.1038/nature03959

Maxam, A. M., and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74, 560–564.

Messacar, K., Parker, S. K., Todd, J. K., and Dominguez, S. R. (2017). Implementation of rapid molecular infectious disease diagnostics: the role of diagnostic and antimicrobial stewardship. *J. Clin. Microbiol.* 55, 715–723. doi: 10.1128/JCM.02264-16

Morganti, S., Tarantino, P., Ferraro, E., D'Amico, P., Viale, G., Trapani, D., et al. (2019). Complexity of genome sequencing and reporting: next generation sequencing (NGS) technologies and implementation of precision medicine in real life. *Crit. Rev. Oncol. Hematol.* 133, 171–182. doi: 10.1016/j.critrevonc.2018.11.008

Mourad, O., Palda, V., and Detsky, A. S. (2003). A comprehensive evidence-based approach to fever of unknown origin. *Arch. Intern. Med.* 163, 545–551.

News in Brief (2020). First Ngs-based COVID-19 diagnostic. *Nat. Biotechnol.* 38:777. doi: 10.1038/s41587-020-0608-y

Nyrén, P., Pettersson, B., and Uhlén, M. (1993). Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem.* 208, 171–175. doi: 10.1006/abio.1993.1024

Pérez-Enciso, M., and Ferretti, L. (2010). Massive parallel sequencing in animal genetics: wherefroms and wheretos. *Anim. Genet.* 41, 561–569. doi: 10.1111/j.1365-2052.2010.02057.x

Petersdorf, R. G., and Beeson, P. B. (1961). Fever of unexplained origin: report on 100 cases. *Medicine* 40, 1–30.

Ren, L.-L., Wang, Y.-M., Wu, Z.-Q., Xiang, Z.-C., Guo, L., Xu, T., et al. (2020). Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin. Med. J.* 133, 1015–1024. doi: 10.1097/CM9.0000000000000722

Reuwer, A. Q., van den Bijllaardt, W., Murk, J. L., Buiting, A. G. M., and Verweij, J. J. (2019). Added diagnostic value of broad-range 16s PCR on periprosthetic tissue and clinical specimens from other normally sterile body sites. *J. Appl. Microbiol.* 126, 661–666. doi: 10.1111/jam.14156

Reyes, A., Carbo, E. C., Harinxma Thoe Slooten, J. S. V., Kraakman, M. E. M., Sidorov, I. A., Claas, E. C. J., et al. (2021). Viral metagenomic sequencing in a cohort of international travellers returning with febrile illness. *J. Clin. Virol.* 143:104940. doi: 10.1016/j.jcv.2021.104940

Ronaghi, M., Uhlén, M., and Nyrén, P. (1998). A sequencing method based on real-time pyrophosphate. *Science* 281, 363–365.

Sanger, F., and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–448.

Shi, J., Yang, N., and Qian, G. (2021). Case report: metagenomic next-generation sequencing in diagnosis of talaromycosis of an immunocompetent patient. *Front. Med.* 8:656194. doi: 10.3389/fmed.2021.656194

Strub, T., Giuliano, S., Ye, T., Bonet, C., Keime, C., Kobi, D., et al. (2011). Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* 30, 2319–2332. doi: 10.1038/onc.2010.612

Tromp, M., Lansdorp, B., Bleeker-Rovers, C. P., Gunnewiek, J. M. K., Kullberg, B. J., and Pickkers, P. (2012). Serial and panel analyses of biomarkers do not improve the prediction of bacteremia compared to one procalcitonin measurement. *J. Infect.* 65, 292–301. doi: 10.1016/j.jinf.2012.06.004

Unger, M., Karanikas, G., Kerschbaumer, A., Winkler, S., and Aletaha, D. (2016). Fever of unknown origin (FUO) revised. *Wien Klin. Wochenschr.* 128, 796–801.

Williams, S. H., Cordey, S., Bhuva, N., Laubscher, F., Hartley, M.-A., Boillat-Blanco, N., et al. (2018). Investigation of the plasma virome from cases of unexplained febrile illness in Tanzania from 2013 to 2014: a comparative analysis between unbiased and vircapseq-vert high-throughput sequencing approaches. *mSphere* 3:e00311-18. doi: 10.1128/mSphere.00311-18

Xiao, M., Liu, X., Ji, J., Li, M., Li, J., Yang, L., et al. (2020). Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 12:57. doi: 10.1186/s13073-020-00751-4

Zhu, H., Zhu, M., Lei, J.-H., Xiao, Y.-L., and Zhao, L.-M. (2021). Metagenomic next-generation sequencing can clinch diagnosis of non-tuberculous mycobacterial infections: a case report. *Front. Med.* 8:679755. doi: 10.3389/fmed.2021.679755

Zou, J., Wang, T., Qiu, T., Chen, Z., Zhou, J., Ma, X., et al. (2022). Clinical characteristics of tuberculous infection following renal transplantation. *Transpl. Immunol.* 70:101523. doi: 10.1016/j.trim.2021.101523

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 09 August 2022
DOI 10.3389/fgene.2022.960731

OPEN ACCESS

EDITED BY
David W. Ussery,
University of Arkansas for Medical Sciences, United States

REVIEWED BY
Chew W. Cheng,
University of Leeds, United Kingdom
Yuri B. Lebedev,
Institute of Bioorganic Chemistry (RAS), Russia

*CORRESPONDENCE
Birte Möhlendick,
birte.moehlendick@uk-essen.de

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology,
a section of the journal Frontiers in Genetics

RECEIVED 03 June 2022
ACCEPTED 15 July 2022
PUBLISHED 09 August 2022

CITATION
Möhlendick B, Schönfelder K, Zacher C, Elsner C, Rohn H, Konik MJ, Thümmler L, Rebmann V, Lindemann M, Jöckel K-H and Siffert W (2022), The GNB3 c.825C>T (rs5443) polymorphism and protection against fatal outcome of corona virus disease 2019 (COVID-19). *Front. Genet.* 13:960731.
doi: 10.3389/fgene.2022.960731

The *GNB3* c.825C>T (rs5443) polymorphism and protection against fatal outcome of corona virus disease 2019 (COVID-19)

Birte Möhlendick[1]*, Kristina Schönfelder[2], Christoph Zacher[1], Carina Elsner[3], Hana Rohn[4], Margarethe J. Konik[4], Laura Thümmler[4,5], Vera Rebmann[5], Monika Lindemann[5], Karl-Heinz Jöckel[6] and Winfried Siffert[1]

[1]Institute of Pharmacogenetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [2]Department of Nephrology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [3]Institute for Virology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [4]Department of Infectious Diseases, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [5]Institute for Transfusion Medicine, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [6]Institute of Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany

Background and aims: Albeit several factors which influence the outcome of corona virus disease (COVID-19) are already known, genetic markers which may predict the outcome of the disease in hospitalized patients are still very sparse. Thus, in this study, we aimed to analyze whether the single-nucleotide polymorphism (SNP) rs5443 in the gene *GNB3*, which was associated with higher T cell responses in previous studies, might be a suitable biomarker to predict T cell responses and the outcome of COVID-19 in a comprehensive German cohort.

Methods: We analyzed the influence of demographics, pre-existing disorders, laboratory parameters at the time of hospitalization, and *GNB3* rs5443 genotype in a comprehensive cohort (N = 1570) on the outcome of COVID-19. In a sub cohort, we analyzed SARS-CoV-2-specific T cell responses and associated *GNB3* rs5443 genotypes. We investigated the influence of all factors on COVID-19 fatality in multivariable analysis.

Results: We found a younger patient age, normotension or absence of diabetes mellitus or cardiovascular diseases, normal blood cell counts, and low inflammatory markers at hospital admission were protective factors against fatal course of disease. In addition, the rs5443 TT genotype was significantly associated with protection against COVID-19 fatality (OR: 0.60, 95% CI: 0.40–0.92, p = 0.02). We also observed significantly increased SARS-CoV-2-specific T cell responses in rs5443 TT genotype carriers (p = 0.01). Although we observed a significant association of the factors described previously in univariate analysis, only a younger age of the patients, normal blood cell counts, and the *GNB3* rs5443 TT genotype remained independent predictors against COVID-19 fatality in multivariable analysis.

Conclusion: Immutable predictors for COVID-19 fatality are relatively rare. In this study we could show that the TT genotype of the SNP rs5443 in the gene *GNB3* is associated with protection against COVID-19 fatality. It was as well correlated to higher SARS-CoV-2-specific T cell responses, which could result in a milder course of disease in those patients. Based on those observations we hereby provide a further prognostic biomarker, which might be used in routine diagnostics as a predictive factor for COVID-19 mortality already upon hospitalization.

KEYWORDS

GNB3, rs5443, genetic association, T cell response, G protein, COVID-19, SARS-CoV-2, disease severity

Introduction

Heterotrimeric guanine-binding proteins (G proteins) transmit signals from the cell surface, trigger intracellular signal cascades, and involve in a wide variety of physiological processes (Klenke et al., 2011). The gene *GNB3* encodes the G protein subunit β3 and is located on chromosome 12p13.31. The ß-subunits are not only the important regulators of the *a*-subunits of G proteins but also intracellular effectors. The synonymous single-nucleotide polymorphism (SNP) rs5443 (c.825C>T; p.S275=) in the gene *GNB3* is associated with several disorders and affects the pharmacodynamics of many different drugs (Klenke et al., 2011). The T allele of this SNP gives rise to the splice variant Gβ3-s, which lacks 123 nucleotides or 41 amino acids. Aberrant splicing results in a dominant gain of function and G protein activation (Siffert et al., 1998). We could show in previous studies that the rs5443 T allele is associated with increased chemotaxis, migration, and proliferation of B lymphoblasts, neutrophils, and T lymphocytes (Virchow et al., 1998; Virchow et al., 1999; Lindemann et al., 2001; Tummala, 2013). Lindemann et al. (2001) could show that $CD4^+$ T cell counts are increased in individuals carrying the rs5443 T allele. Therefore, it appears that individuals carrying the T allele show an increased function of their cellular immune system.

Adaptive immune responses, especially those of the T cells, are of major importance in SARS-CoV-2 infection. Virus-specific $CD4^+$ and $CD8^+$ T cells produce effector cytokines and exert cytotoxic activity in most patients with SARS-CoV-2 infection, whereas neutralizing antibodies directly interfere with viral entry of host cells (Jung and Shin, 2021). Nevertheless, patients with corona virus disease 2019 (COVID-19) not only show lower proportions of SARS-CoV-2-specific $CD4^+$ or $CD8^+$ T cells but also B cells and NK cells, with increasing disease severity (Huang et al., 2020; Peng et al., 2020; Zeng et al., 2020; Olea et al., 2021). Zeng et al. (2020) observed $CD4^+$ T cell lymphopenia in all severe and fatal cases with SARS-CoV-2 infection in their study. Furthermore, the authors could show that prolonged activation and exhaustion of $CD8^+$ T cells were associated with COVID-19 severity. In single-cell transcriptomic analyses, encompassing over 80,000 virus-reactive $CD8^+$ T single cells, Kusnadi et al. (2021) could show that SARS-CoV-2-reactive $CD8^+$ cells exhibited exhausted phenotypes with a decreased capacity to produce cytokines in severely ill COVID-19 patients.

In light of these observations, we hypothesized that the SNP rs5443 in the gene *GNB3* might influence the T cell response in COVID-19 patients as well and, thereby, the outcome of the disease. To answer this question we analyzed the SNP rs5443 in the gene *GNB3* in a comprehensive retrospective German cohort with SARS-CoV-2 infection and its influence upon T cell response and course of COVID-19.

Methods

Study participants, recruitment, and outcome of the patients

The study was conducted following the approval of the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen (20-9230-BO) and in cooperation with the West German Biobank (WBE; 20-WBE-088). Written informed consent was obtained from the study patients.

Enrollment started on 11 March 2020, and ended on 18 May 2021. Altogether, 1,570 SARS-CoV-2-positive patients with at least one positive real-time reverse transcription polymerase chain reaction (RT-PCR) test result were consecutively recruited for the study. Follow-up was completed on 30 June 2021, and at that time all patients either were discharged from the hospital as "cured" or had a fatal outcome of the disease. The clinical outcome was defined as follows according to the criteria of the ECDC (European Center of Disease Prevention and Control, 2021)—"mild': outpatients (N = 205); "hospitalized": inpatients (N = 760); "severe": hospitalized patients admitted to an intensive care unit and/or became dependent on mechanical ventilation (N = 292); "fatal" all cases of COVID-19-related deaths during the hospital stay or within a follow-up of 30 days (N = 313). In contrast to the ECDC classification, where patients counted up to three times, every patient counted only once, according to the worst clinical outcome

observed during the hospital stay in our study. The patients included in this study were of Caucasian origin.

For further statistical analyses, demographic data, medical history, and hematological parameters (erythrocyte, platelet, neutrophil, and lymphocyte counts) at the time of hospital admission were documented for each patient. The medical history included pre-existing disorders of the cardiovascular system (e.g., myocardial infarction, coronary heart disease but not arterial hypertension), arterial hypertension, and diabetes mellitus.

Neutrophil–lymphocyte ratio, platelet–lymphocyte ratio, and systemic immune-inflammation index were calculated as inflammatory markers. The neutrophil-lymphocyte ratio (NLR) is calculated by dividing the number of neutrophils per nanoliter (nl) by the number of lymphocytes per nl from a peripheral blood sample. Similarly, the platelet-lymphocyte ratio (PLR) is calculated, where the number of platelets per nl is divided by the number of lymphocytes per nl in a peripheral blood sample. For the systemic immune-inflammation index (SII), the platelet counts per nl were multiplied by the number of neutrophils per nl and then divided by lymphocyte counts per nl in a peripheral blood sample.

Interferon-γ ELISpot assay

SARS-CoV-2-specific T cell responses were analyzed in 182 randomly selected SARS-CoV-2-positive patients using interferon-γ (IFN-γ) ELISpot assays as previously described (Schwarzkopf et al., 2021). Briefly, ELISpot stripes containing polyvinylidene difluoride (PVDF) membranes (MilliporeSigma™ MultiScreen™ HTS, Fisher Scientific, Schwerte, Germany) were activated with 50 µl of 35% ethanol for 10 s and washed with distilled water. Plates were then coated for 3 hours with 60 µl of monoclonal antibodies against IFN-γ (10 µg/ml of clone 1-D1K, Mabtech, Nacka, Sweden). Thereafter, ELISpot plates were washed and then blocked with 150 µl AIM-V® (Thermo Scientific, Dreireich, Germany). After 30 min at 37°C, AIM-V® was discarded, and duplicates of 250,000 peripheral blood mononuclear cells (PBMC) were grown in the presence or absence of either PepTivator® SARS-CoV-2 protein S1/S2 (600 pmol/ml, Miltenyi Biotec, Bergisch Gladbach, Germany) in 150 µl of AIM-V®. The peptide mix of the S1/S2 protein consists mainly of 15-mer sequences with 11 amino acids overlap, covering the immunodominant sequence domains of the surface glycoprotein of SARS-CoV-2. After 19 h of incubation at 37°C, the ELISpot plates were washed, and captured IFN-γ was detected by incubation for 1 hour with 50 µl of the alkaline phosphatase-conjugated monoclonal antibody against IFN-γ (clone 7-B6-1, Mabtech, Stockholm, Sweden), diluted 1:200 with PBS plus 0.5% bovine serum albumin (BSA). After further washing, 50 µl of nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl-phosphate (NBT/BCIP) was added, and purple spots appeared within 7 min. Spot numbers were analyzed by an ELISpot reader (AID Fluorospot, Autoimmun Diagnostika GmbH, Strassberg, Germany). Mean values of duplicate cell cultures were considered. We determined SARS-CoV-2-specific spots by spot increment, defined as stimulated minus non-stimulated values. Stimulated spot numbers > 3-fold higher than negative (unstimulated) controls combined with an increment value of >3 to the antigen were considered positive. Of note, the negative controls reached a mean value of less than one spot.

Genotyping of *GNB3* rs5443 (c.825C>T)

Genomic DNA was extracted from 200 µl EDTA-blood using the QIAamp® DNA Blood Mini Kit (Qiagen, Hilden, Germany). Polymerase chain reaction (PCR) was performed with 2 µl genomic DNA and 30 µl *Taq* DNA-Polymerase 2x Master Mix Red (Ampliqon, Odense, Denmark), with the following conditions: initial denaturation 94°C for 3 min; 35 cycles with denaturation 94°C for 30 s, annealing at 66°C for 30 s, and elongation 72°C for 30 s each; final elongation 72°C for 10 min (forward primer: 5′ GCT GCC CAG GTC TGA TCC C 3' and reverse primer 3′ TGG GGA GGG TCC TTC CAG C 5′). PCR products were digested with *Bse*DI (Thermo Scientific, Dreireich, Germany), and restriction fragments were analyzed by agarose gel electrophoresis. The various genotype results from restriction fragment length polymorphism (RFLP)-PCR were validated by Sanger sequencing.

Statistical analyses

Correlation of demographics (sex and medical history) and outcome of COVID-19 were calculated using Pearson's chi square (χ^2) statistics using the Baptista–Pike method for the odds ratio (OR) and 95% confidence interval (CI). One-way analysis of variance (ANOVA) was performed using the Kruskal–Wallis test with Dunn's multiple comparison to assess the influence of age, hematological parameters, or inflammatory markers on COVID-19 severity. To calculate thresholds for the laboratory values, which correlate with fatal course of disease receiver operating characteristic (ROC) analysis, Youden`s J statistic was performed.

The number of patients with fatal outcome of disease, for whom IFN-γ ELISpot analyses could be performed, was relatively small. Thus, we defined additional groups to perform statistical analyses to estimate the influence of the T cell response on COVID-19 severity in our cohort. Therefore, patients from the categories "mild" and "hospitalized" were grouped together to the group "moderate," whereas the patients with "severe" and "fatal" COVID-19 were consolidated to the group "serious." The differences in T cell responses as analyzed by IFN-γ ELISpot between patients with "moderate" and "serious" COVID-19 was estimated by Mann–Whitney test.

TABLE 1 Demographics, clinical characteristics, and outcome of the disease in SARS-CoV-2-positive patients. Classification according to the COVID-19 surveillance report of the ECDC: category "mild" is a case that has not been reported as hospitalized or dead. A "severe" case has been admitted to intensive care and/or required mechanical respiratory support. All values are given in medians and interquartile ranges (IQR), except from sex and medical history, which are reported in absolute counts and percentages.

Characteristics	All patients (N = 1570)	Mild (N = 205)	Hospitalized (N = 760)	Severe (N = 292)	Fatal (N = 313)	p-value
Age–years	62.0 (49.0–76.0)	47.0 (34.5–64.0)	62.0 (48.3–76.0)	59.0 (50.0–70.0)	71.0 (59.5–82.0)	$p < 0.0001$
Male sex	910 (58.0)	107 (52.2)	416 (54.7)	185 (63.4)	202 (64.5)	$p = 0.002$
Medical history						
Cardiovascular system[a]	547 (34.8)	11 (5.4)	257 (33.8)	111 (38.0)	168 (53.7)	$p < 0.0001$
Arterial hypertension	748 (47.6)	29 (14.1)	373 (49.1)	149 (51.0)	197 (62.9)	$p < 0.0001$
Diabetes mellitus	404 (25.7)	14 (6.8)	214 (28.2)	76 (26.0)	100 (31.9)	$p = 0.001$
Hematological parameters						
Erythrocytes/nl	4.4 (3.8–4.8)	4.6 (4.2–4.9)	4.4 (4.0–4.9)	4.4 (3.8–4.8)	4.0 (3.4–4.6)	$p < 0.0001$
Platelets/nl	202.0 (156.0–260.0)	204.0 (164.0–270.5)	205.0 (157.0–255.0)	209.0 (169.0–292.0)	189.0 (135.0–242.0)	$p < 0.0001$
Neutrophils/nl	4.9 (3.1–7.5)	3.7 (2.7–5.1)	3.9 (2.6–5.8)	6.3 (4.2–9.3)	7.7 (5.2–11.7)	$p < 0.0001$
Lymphocytes/nl	0.9 (0.7–1.3)	1.1 (0.9–1.5)	1.0 (0.7–1.4)	0.8 (0.6–1.1)	0.7 (0.5–1.1)	$p < 0.0001$
Inflammatory markers						
NLR	5.0 (2.9–9.9)	3.1 (2.1–4.7)	3.8 (2.4–6.2)	7.9 (4.5–13.0)	11.1 (6.0–18.5)	$p < 0.0001$
PLR	217.8 (151.1–326.7)	176.3 (139.1–268.9)	197.7 (140.7–285.5)	269.8 (181.4–418.4)	252.6 (162.9–414.2)	$p < 0.0001$
SII	1031.0 (523.9–2206.0)	717.2 (385.7–1055.0)	769.6 (399.8–1417.0)	1680 (921.9–3466.0)	1917.0 (1010.0–4019.0)	$p < 0.0001$

[a]Cardiovascular system: for example, myocardial infarction, coronary heart disease but not arterial hypertension. Abbreviations: nl = nanoliter; NLR, neutrophil–lymphocyte ratio; PLR, platelet–lymphocyte ratio; SII, systemic immune-inflammation index.

Hardy–Weinberg equilibrium (HWE) was calculated using Pearson's chi square (χ^2) goodness of fit test, and samples were considered as deviant from HWE at a significance level of $p < 0.05$.

For genetic association, we calculated OR and 95% CI by Pearson's chi square (χ^2) statistics using the Baptista–Pike method for OR and 95% CI, respectively. p-values are reported two-sided, and values of <0.05 were considered significant. One-way analysis of variance (ANOVA) was performed using Kruskal–Wallis test with Dunn's multiple comparison test to determine the influence of *GNB3* rs5443 genotype on T cell response as measured by IFN-γ ELISpot assay.

Multivariable analysis was performed to estimate independency of the variables age, sex, medical history, laboratory parameters, and *GNB3* rs5443 genotypes by stepwise Cox regression (likelihood ratio test, backward).

Results

From 11 March 2020 to 30 June 2021, we enrolled and studied 1,570 SARS-CoV-2-positive patients to determine the association of the SNP rs5443 in the gene *GNB3*, with severity of COVID-19. In a sub group of patients (N = 182), who were representative for all severity groups, we additionally analyzed the T cell response to SARS-CoV-2-specific antigens. The demographics and clinical characteristics of the patients are summarized in Table 1. We observed that about 20% of all patients (inpatients and outpatients) and 23% of the hospitalized patients had a fatal outcome of COVID-19. With increasing severity of the disease, we found significantly more elderly and male patients and those who had arterial hypertension, cardiovascular disorders, or diabetes mellitus as pre-existing medical disorders (Table 1). The number of platelets, erythroctytes, and lymphocytes decreased significantly, whereas the neutrophil counts increased with disease severity ($p < 0.0001$, ANOVA). Regarding the inflammatory markers, NLR, PLR, and SII, we observed significantly higher values with increasing severity of COVID-19 as well.

SARS-CoV-2-specific T cell response and *GNB3* rs5443 genotype

In 182 patients, we performed IFN-γ ELISpot assays to determine T cell response to SARS-CoV-2-specific antigens. We were able to analyze patients from all severity groups: "mild" (N = 79); "hospitalized" (N = 82); "severe" (N = 17), and "fatal" (N = 4). The number of patients with fatal outcome of disease, for whom IFN-γ ELISpot analyses could be performed, was relatively small. Thus, we defined additional groups to perform statistical analyses to estimate the influence of the T cell response on COVID-19 severity in

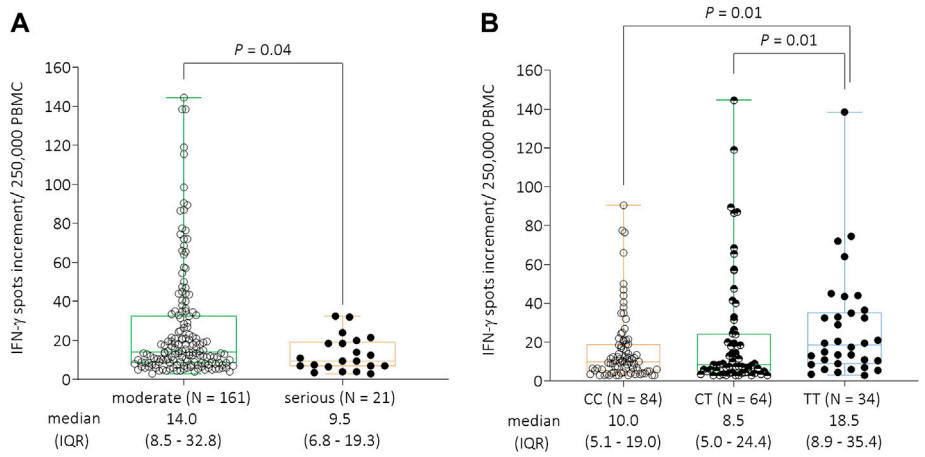


FIGURE 1
(A) IFN-γ ELISpot responses to S1/S2 protein of SARS-CoV-2 per 250,000 peripheral blood mononuclear cells stratified by COVID-19 severity. Due to the low number of cases in the individual groups, patients from the categories "mild" and "hospitalized" were grouped together to the group "moderate," whereas the patients with "severe" and "fatal" COVID-19 were consolidated to the group "serious." There was a significant decline of IFN-γ spots increment comparing the "serious" (N = 21, median = 9.5, and IQR = 6.8–19.3) to the "moderate" (N = 161, median = 13.5, and IQR = 8.0–35.0) groups (p = 0.04). **(B)** IFN-γ ELISpot responses to S1/S2 protein of SARS-CoV-2 per 250,000 peripheral blood mononuclear cells stratified by the *GNB3* rs5443 genotype. Individuals with the TT genotype had significantly higher spots increment (median = 18.5, and IQR = 8.9–35.4) compared to CT (median = 8.5, IQR = 5.0–24.4, and p = 0.01) or CC genotype carriers (median = 10.0, IQR = 5.1–19.0, and p = 0.01). Abbreviations: IFN-γ = interferon gamma; PBMC = peripheral blood mononuclear cells; IQR = interquartile range.

TABLE 2 *GNB3* rs5443 (c.825C>T) genotype distribution among all patients with SARS-CoV-2 infection and subdivided according to the severity of COVID-19.

	All patients (N = 1,570)	Mild (N = 205)	Hospitalized (N = 760)	Severe (N = 292)	Fatal (N = 313)
GNB3 rs5443 CC	700 (44.6)	89 (43.4)	330 (43.4)	121 (41.4)	160 (51.1)
GNB3 rs5443 CT	666 (42.4)	90 (43.9)	324 (42.6)	130 (44.5)	122 (39.0)
GNB3 rs5443 TT	204 (13.0)	26 (12.7)	106 (13.9)	41 (14.0)	31 (9.9)
Minor allele frequency (T)	0.34	0.35	0.35	0.36	0.29

our cohort. Therefore, patients from the categories "mild" and "hospitalized" were grouped together to the group "moderate," whereas the patients with "severe" and "fatal" COVID-19 were consolidated to the group "serious." We observed a significant decline of spots increment in the IFN-γ ELISpot assay comparing the "serious" group (N = 21, median = 9.5, and IQR = 6.8–19.3) to the "moderate" group (N = 161, median = 14.0, IQR = 8.5–32.8, p = 0.04, Figure 1A).

In a next step, we analyzed the influence of *GNB3* rs5443 genotypes on IFN-γ production against SARS-CoV-2-specific antigens. Here, we found a significant increase of IFN-γ spots increment in TT genotype carriers (median = 18.5 and IQR = 8.9–35.4) compared to those with CC genotype (median = 10.0, and IQR = 5.1–19.0) or CT genotype (median = 8.5, and IQR = 5.0–24.4) (both p = 0.01, respectively, Figure 1B).

GNB3 rs5443 as a protective factor against COVID-19 fatality

Overall, the observed genotypes for *GNB3* rs5443 were compatible with HWE in patients with "mild" (p = 0.66), "hospitalized" (p = 0.07), "severe" (p = 0.52), and "fatal" (p = 0.28) SARS-CoV-2 infection. Genotype distributions for all patients and the different groups according to severity of SARS-CoV-2 infection are shown in Table 2. Notably, we observed very similar rs5443 T allele frequencies (35%–36%) in all groups, except from those patients with a "fatal" outcome of COVID-19 (29%). Thus, we estimated, whether T allele or TT genotype carriers might be protected more effectively against fatal outcome of the disease. We found a significant association for protection against COVID-19 fatality in rs5443 TT genotype carriers comparing all patients ("mild," "hospitalized," and "severe") with SARS-CoV-2 infection

TABLE 3 Protective factors against COVID-19 fatality. Abbreviations: nl = nanoliter; NLR = neutrophil-lymphocyte ratio; PLR = platelet-lymphocyte ratio; SII = systemic immune-inflammation index; OR = odds ratio; CI = confidence interval, NS = not significant in stepwise multivariable analysis.

	Univariate analysis		Multivariable analysis	
Factor	**OR (95% CI)**	***p*-value**	**Or ([95% CI)**	***p*-value**
Age (<62 years)	0.35 (0.27–0.45)	<0.0001	0.47 [0.34–0.64)	<0.0001
Sex (female)	0.71 (0.55–0.92)	0.01	NS	NS
Absence of				
Diseases of cardiovascular system	0.41 (0.32–0.53)	<0.0001	NS	NS
Arterial hypertension	0.53 (0.41–0.68)	<0.0001	NS	NS
Diabetes mellitus	0.75 (0.57–0.98)	0.04	NS	NS
Hematological parameters				
Erythrocytes (≥4.0/nl)	0.27 (0.21–0.34)	<0.0001	0.70 (0.52–0.94)	0.02
Platelets (≥133.5/nl)	0.40 (0.29–0.54)	<0.0001	0.42 (0.30–0.60)	<0.0001
Neutrophils (≤6.6/nl)	0.28 (0.21–0.36)	<0.0001	0.32 (0.23–0.45)	<0.0001
Lymphocytes (≥0.9/nl)	0.41 (0.32–0.53)	<0.0001	0.55 (0.41–0.74)	<0.0001
Inflammatory markers				
NLR (<7.3)	0.24 (0.18–0.31)	<0.0001	NS	NS
PLR (<224.6)	0.60 (0.46–0.78)	<0.0001	NS	NS
SII (<1206.4)	0.32 (0.24–0.42)	<0.0001	NS	NS
GNB3 rs5443 TT genotype	0.60 (0.40–0.92)	0.02	0.65 (0.44–0.96)	0.03

and with those who died from COVID-19 (OR: 0.60, 95% CI: 0.40–0.92; $p = 0.02$, Table 3).

Thereupon, we performed multivariable analysis to analyze the independence of the *GNB3* rs5443 TT genotype in comparison to the other predictive parameters: age, pre-existing disorders, hematological parameters, and inflammatory markers. We performed ROC analysis and Youden`s statistic for the numeric variables to estimate a threshold above which the risk for COVID-19 fatality significantly decreased. We found that a younger patient age (<62 years; $p < 0.0001$), erythrocyte (≥4.0/nl; $p = 0.02$), platelet (≥133.5/nl; $p < 0.0001$), neutrophil (<6.6/nl; $p < 0.0001$), and lymphocyte (≥0.9/nl; $p < 0.0001$) counts above these respective thresholds at the time of admission to hospital, and the *GNB3* rs5443 TT genotype ($p = 0.03$) remained independent predictors for protection against COVID-19 fatality (Table 3).

Discussion

Remarkably, we observed that the TT genotype of the SNP rs5443 in the gene *GNB3* was associated with a higher T cell response as estimated by IFN-γ ELISpot assay in our patients. We could not find an association of *GNB3* genotype to lymphocyte or T cell counts. Thus, it seems that the increased T cell response in TT genotype carriers might be related to an increased activation of T cells. Early development of $CD8^+$ T cell responses is correlated to a more effective viral clearance and a mild course of COVID-19. Patients with severe disease display early onset of inflammation as well as delayed and relatively excessive adaptive immune response (Moss, 2022). The SNP rs5443 in the gene *GNB3* was not only correlated to higher T cell responses but also to a significantly reduced risk for COVID-19 fatality in our study in univariate and multivariable analyses.

The underlying mechanism of the influence of *GNB3* genotype on T cell response remains elusive. Juno (2014) could show that *GNB3* TT genotype carriers had a significantly lower *LAG-3* gene expression. The *LAG-3* (lymphocyte activation gene 3) gene is localized on chromosome 12 nearby to *GNB3*, nevertheless there are no SNPS in the gene *LAG-3* in high linkage disequilibrium with rs5443, which could be causative for the different gene expression in *GNB3* TT genotype carriers. LAG-3 was found to be expressed on dysfunctional or exhausted T cells in chronic viral infections and correlated with severity of the infection (Blackburn et al., 2009; Richter et al., 2010). Further studies are needed to analyze whether a reduced *LAG-3* expression is responsible for the T cell activation in *GNB3* TT genotype carriers.

We found that the comorbidities arterial hypertension, other disorders of the cardiovascular system and diabetes mellitus were associated with COVID-19 fatality in univariate analysis. This has already been extensively described in a multitude of studies and meta-analyses (Zhou et al., 2020). A variety of other factors, for example, age, sex, or laboratory parameters, have also been identified to influence the course of COVID-19 (Hobohm et al., 2022). The

infection-fatality ratio of COVID-19 significantly increases through ages 30, 60, and 90 years (COVID-19 Forecasting Team, 2022). Thus, we observed that a younger age (<62 years) was an independent protective factor against COVID-19 fatality in our study as well. Nevertheless, we could not confirm the independent influence of other consistent factors, like sex or pre-existing disorders, in a multivariable analysis.

Normal cell counts of lymphocytes and platelets upon hospital admission are associated with a significantly reduced risk for fatal outcome of COVID-19. Impaired adaptive immune responses as reflected by low counts of white blood cells together with augmented inflammation serve as a good predictor for the course of the disease (Qin et al., 2022). In our study, we noticed impaired white blood cell counts in individuals with severe COVID-19 as well. Several studies could show that the inflammatory markers NLR, PLR, and SII determined upon hospital admission are good predictive markers for in-hospital mortality (Fois et al., 2020; Wang et al., 2021; Sarkar et al., 2022). We also found a significant association for COVID-19 fatality and high NLR, PLR, and SII in the univariate analysis in our study. Nonetheless, those markers did not reach statistical significance in the multivariable analysis. Therefore, it seems even more important to find persistent markers that can predict the course of COVID-19 disease.

Together with a younger patient age, a normal white blood cell count at hospital admission, the *GNB3* rs5443 TT genotype remained an independent protective factor against COVID-19 fatality in our study. Immutable predictors are still relatively rare, thus analyses of genetic host factors might be useful in predicting severity, which could be implemented in routine diagnostics.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials; further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Medical Faculty of the University of Duisburg-Essen. The patients/participants provided their written informed consent to participate in this study.

Author Contributions

BM: conceptualization, resources, methodology, formal analysis, investigation, supervision, data curation, funding acquisition, project administration, visualization, writing—original draft, and writing—review and editing. KS: resources, methodology, data curation, formal analysis, validation, and writing—review, and editing. CZ, CE, HR, MK, LT, and VR: data curation, resources, and investigation. ML: validation and writing—review and editing. WS and K-HL: conceptualization, validation, supervision, and writing—review, and editing.

Funding

This work was supported by a grant (to BM) from the Stiftung Universitätsmedizin Essen of the Medical Faculty Essen. The funder of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in decision of submitting the article for publication. We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen.

Acknowledgments

We kindly thank Iris Manthey, Grit Müller, and Stephanie Büscher for technical support (Institute of Pharmacogenetics). We thank Martina Wachsmann and Petra Kasper for their support with the samples from the central laboratory (University Hospital Essen).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Blackburn, S. D., Shin, H., Haining, W. N., Zou, T., Workman, C. J., Polley, A., et al. (2009). Coregulation of CD8+ T cell exhaustion by multiple inhibitory receptors during chronic viral infection. *Nat. Immunol.* 10, 29–37. doi:10.1038/ni.1679

European Center of Disease Prevention and Control. (2021) ECDC surveillance report. Available at: https://covid19-surveillance-report.ecdc.europa.eu/. [Accessed September 10, 2021],

Fois, A. G., Paliogiannis, P., Scano, V., Cau, S., Babudieri, S., Perra, R., et al. (2020). The systemic inflammation index on admission predicts in-hospital mortality in COVID-19 patients. *Molecules* 25, E5725. doi:10.3390/molecules25235725

Forecasting Team (2022). Variation in the COVID-19 infection-fatality ratio by age, time, and geography during the pre-vaccine era: A systematic analysis. *Lancet* 399, 1469–1488. doi:10.1016/S0140-6736(21)02867-1

Hobohm, L., Sagoschen, I., Barco, S., Schmidtmann, I., Espinola-Klein, C., Konstantinides, S., et al. (2022). Trends and risk factors of in-hospital mortality of patients with COVID-19 in Germany: Results of a large nationwide inpatient sample. *Viruses* 14, 275. doi:10.3390/v14020275

Huang, W., Berube, J., McNamara, M., Saksena, S., Hartman, M., Arshad, T., et al. (2020). Lymphocyte subset counts in COVID-19 patients: A meta-analysis. *Cytom. A* 97, 772–776. doi:10.1002/cyto.a.24172

Jung, M. K., and Shin, E.-C. (2021). Phenotypes and functions of SARS-CoV-2-reactive T cells. *Mol. Cells* 44, 401–407. doi:10.14348/molcells.2021.0079

Juno, J. (2014). "Contribution of guanine nucleotide binding protein beta polypeptide 3 (GNB3) and lymphocyte activation gene 3 (LAG-3) to HIV susceptibility and immune dysfunction," (Winnipeg: University of Manitoba, Department of Medical Microbiology). Thesis.

Klenke, S., Kussmann, M., and Siffert, W. (2011). The GNB3 C825T polymorphism as a pharmacogenetic marker in the treatment of hypertension, obesity, and depression. *Pharmacogenet. Genomics* 21, 594–606. doi:10.1097/FPC.0b013e3283491153

Kusnadi, A., Ramírez-Suástegui, C., Fajardo, V., Chee, S. J., Meckiff, B. J., Simon, H., et al. (2021). Severely ill COVID-19 patients display impaired exhaustion features in SARS-CoV-2-reactive CD8+ T cells. *Sci. Immunol.* 6, eabe4782. doi:10.1126/sciimmunol.abe4782

Lindemann, M., Virchow, S., Ramann, F., Barsegian, V., Kreuzfelder, E., Siffert, W., et al. (2001). The G protein beta3 subunit 825T allele is a genetic marker for enhanced T cell response. *FEBS Lett.* 495, 82–86. doi:10.1016/s0014-5793(01)02339-0

Moss, P. (2022). The T cell immune response against SARS-CoV-2. *Nat. Immunol.* 23, 186–193. doi:10.1038/s41590-021-01122-w

Olea, B., Albert, E., Torres, I., Amat, P., Remigia, M. J., Gozalbo-Rovira, R., et al. (2021). Adaptive immune responses to SARS-CoV-2 in recovered severe COVID-19 patients. *J. Clin. Virol.* 142, 104943. doi:10.1016/j.jcv.2021.104943

Peng, Y., Mentzer, A. J., Liu, G., Yao, X., Yin, Z., Dong, D., et al. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* 21, 1336–1345. doi:10.1038/s41590-020-0782-6

Qin, R., He, L., Yang, Z., Jia, N., Chen, R., Xie, J., et al. (2022). Identification of parameters representative of immune dysfunction in patients with severe and fatal COVID-19 infection: A systematic review and meta-analysis. *Clin. Rev. Allergy Immunol.*, 1–33. doi:10.1007/s12016-021-08908-8

Richter, K., Agnellini, P., and Oxenius, A. (2010). On the role of the inhibitory receptor LAG-3 in acute and chronic LCMV infection. *Int. Immunol.* 22, 13–23. doi:10.1093/intimm/dxp107

Sarkar, S., Kannan, S., Khanna, P., and Singh, A. K. (2022). Role of platelet-to-lymphocyte count ratio (PLR), as a prognostic indicator in COVID-19: A systematic review and meta-analysis. *J. Med. Virol.* 94, 211–221. doi:10.1002/jmv.27297

Schwarzkopf, S., Krawczyk, A., Knop, D., Klump, H., Heinold, A., Heinemann, F. M., et al. (2021). Cellular immunity in COVID-19 convalescents with PCR-confirmed infection but with undetectable SARS-CoV-2-specific IgG. *Emerg. Infect. Dis.* 27, 122–129. doi:10.3201/2701.203772

Siffert, W., Rosskopf, D., Siffert, G., Busch, S., Moritz, A., Erbel, R., et al. (1998). Association of a human G-protein beta3 subunit variant with hypertension. *Nat. Genet.* 18, 45–48. doi:10.1038/ng0198-45

Tummala, H. (2013). The alternate GNB3 splice variant, Gβ3s, exhibits an altered signalling response to EGF stimulation, which leads to enhanced cell migration. *Biodiscovery* 9, e893. doi:10.7750/BioDiscovery.2013.9.3

Virchow, S., Ansorge, N., Rosskopf, D., Rübben, H., and Siffert, W. (1999). The G protein beta3 subunit splice variant Gbeta3-s causes enhanced chemotaxis of human neutrophils in response to interleukin-8. *Naunyn. Schmiedeb. Arch. Pharmacol.* 360, 27–32. doi:10.1007/s002109900040

Virchow, S., Ansorge, N., Rübben, H., Siffert, G., and Siffert, W. (1998). Enhanced fMLP-stimulated chemotaxis in human neutrophils from individuals carrying the G protein beta3 subunit 825 T-allele. *FEBS Lett.* 436, 155–158. doi:10.1016/s0014-5793(98)01110-7

Wang, Y., Zhao, J., Yang, L., Hu, J., and Yao, Y. (2021). Value of the neutrophil-lymphocyte ratio in predicting COVID-19 severity: A meta-analysis. *Dis. Markers* 2021, 2571912. doi:10.1155/2021/2571912

Zeng, Q., Li, Y.-Z., Dong, S.-Y., Chen, Z.-T., Gao, X.-Y., Zhang, H., et al. (2020). Dynamic SARS-CoV-2-specific immunity in critically ill patients with hypertension. *Front. Immunol.* 11, 596684. doi:10.3389/fimmu.2020.596684

Zhou, Y., Yang, Q., Chi, J., Dong, B., Lv, W., Shen, L., et al. (2020). Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: A systematic review and meta-analysis. *Int. J. Infect. Dis.* 99, 47–56. doi:10.1016/j.ijid.2020.07.029

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 29 August 2022
DOI 10.3389/fgene.2022.932043

OPEN ACCESS

EDITED BY
Kouki Matsuda,
National Center For Global Health and Medicine, Japan

REVIEWED BY
Dinesh Mohanraj,
The University of Manchester, United Kingdom
Shiang-Jong Tzeng,
National Taiwan University, Taiwan

*CORRESPONDENCE
Ieva Čiučiulkaitė,
ieva.ciuciulkaite@uk-essen.de

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Genetics

RECEIVED 29 April 2022
ACCEPTED 01 August 2022
PUBLISHED 29 August 2022

CITATION
Čiučiulkaitė I, Möhlendick B, Thümmler L, Fisenkci N, Elsner C, Dittmer U, Siffert W and Lindemann M (2022), GNB3 c.825c>T polymorphism influences T-cell but not antibody response following vaccination with the mRNA-1273 vaccine.
Front. Genet. 13:932043.
doi: 10.3389/fgene.2022.932043

GNB3 c.825c>T polymorphism influences T-cell but not antibody response following vaccination with the mRNA-1273 vaccine

Ieva Čiučiulkaitė[1]*, Birte Möhlendick[1], Laura Thümmler[2], Neslinur Fisenkci[2], Carina Elsner[3], Ulf Dittmer[3], Winfried Siffert[1] and Monika Lindemann[2]

[1]Institute of Pharmacogenetics, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [2]Institute for Transfusion Medicine, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, [3]Institute for Virology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany

Background: Immune responses following vaccination against COVID-19 with different vaccines and the waning of immunity vary within the population. Genetic host factors are likely to contribute to this variability. However, to the best of our knowledge, no study on G protein polymorphisms and vaccination responses against COVID-19 has been published so far.

Methods: Antibodies against the SARS-CoV-2 spike protein and T-cell responses against a peptide pool of SARS-CoV-2 S1 proteins were measured 1 and 6 months after the second vaccination with mRNA-1273 in the main study group of 204 participants. Additionally, antibodies against the SARS-CoV-2 spike protein were measured in a group of 597 participants 1 month after the second vaccination with mRNA-1273. Genotypes of *GNB3* c.825C>T were determined in all participants.

Results: The median antibody titer against the SARS-CoV-2 spike protein and median values of spots increment in the SARS-CoV-2 IFN-γ ELISpot assay against the S1-peptide pool were significantly decreased from months 1 to 6 ($p < 0.0001$). Genotypes of *GNB3* c.825C>T had no influence on the humoral immune response. At month 1, CC genotype carriers had significantly increased T-cell responses compared to CT ($p = 0.005$) or TT ($p = 0.02$) genotypes. CC genotype carriers had an almost 6-fold increased probability compared to TT genotype carriers and an almost 3-fold increased probability compared to T-allele carriers to mount a SARS-CoV-2-specific T-cell response above the median value.

Conclusion: CC genotype carriers of the *GNB3* c.825C>T polymorphism have an increased T-cell immune response to SARS-CoV-2, which may indicate better T-cell-mediated protection against COVID-19 after vaccination with mRNA-1273.

KEYWORDS

GNB3 c.825C>T, COVID-19, SARS-CoV-2, mRNA-1273, antigen-specific T-cell response, ELISpot, SARS-CoV-2 spike antibody titer

1 Introduction

Antibodies and T-cells play an important role in both the outcome of COVID-19 and vaccination against it. Interaction between the angiotensin-converting enzyme 2 receptor expressed on the host cells and the receptor-binding domain in the spike (S) 1 subunit of the SARS-CoV-2 spike protein allows the virus to enter the host cell (Harrison et al., 2020). Vaccines against COVID-19 encode this SARS-CoV-2 spike protein and induce an immune response (Martinez-Flores et al., 2021). Immune responses following vaccination against COVID-19 with different vaccines and the waning of immunity vary within the population (Collier et al., 2021). Common factors such as age, sex, pre-existing conditions, or immunosuppressive therapy have been investigated and shown to contribute to this variability (Geisen et al., 2021; Lindemann et al., 2021; Simon et al., 2021; Steensels et al., 2021; Widge et al., 2021). In addition, genetic host factors are also likely to contribute to this variability (Crocchiolo et al., 2022; Gutierrez-Bautista et al., 2022). However, to the best of our knowledge, no studies on vaccination against COVID-19 and G protein polymorphisms have been published so far.

Here, we investigated whether genotypes of the c.825C>T polymorphism in the gene *GNB3* (rs5443) may influence the immune response after vaccination against COVID-19. This polymorphism exerts diverse influences on G protein-mediated signaling by generating a splice variant of the G protein subunit beta-3 (Siffert et al., 1998). Previous studies have shown that the *GNB3* c.825C>T polymorphism affects the immune response after stimulation with various recall antigens and after vaccination against the hepatitis B virus (HBV) (Lindemann et al., 2001; Lindemann et al., 2002).

2 Methods

2.1 Study group

For the study, 2,526 healthcare workers from the University Hospital Essen (Essen, Germany) were recruited. From this study cohort, we gathered a homogeneous group of 204 participants aged between 18–40 years for further investigations. All participants in this study group were non-obese, non-smokers, and were healthy or had minor health issues, but no immunosuppressive conditions or cardiovascular diseases. Immune responses after the vaccination in the study group did correlate neither with age nor with BMI. Furthermore, there were no differences in immune responses between healthy participants and participants with minor health issues. The allele frequencies of *GNB3* c.825C>T are differently distributed in African and East Asian populations. In this study, merely two participants belonged to these populations and they constituted less than 1% of our study group. The selection was based on questionnaires and the flow chart of enrollment is shown in Supplementary Figure S1. For additional investigations of antibody titers, we established an age-matched replication group of 597 participants. All participants in both study groups were vaccinated twice with the COVID-19 vaccine mRNA-1273 (Moderna Inc.). None of the participants had a history of SARS-CoV-2 infection and all tested negative for antibodies against the SARS-CoV-2 nucleocapsid protein. The investigations were reviewed and approved by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen (21–10005-BO). All participants provided their written informed consent to participate in this study.

2.2 Study design

Blood samples were taken from all participants 1 and 6 months after the second vaccination with mRNA-1273. We measured antibody titers against the SARS-CoV-2 S protein and the SARS-CoV-2 nucleocapsid protein and determined genotypes of the *GNB3* c.825C>T polymorphism. In addition, in the main study group of 204 participants, the T-cell response against the S1 peptide pool was measured using the SARS-CoV-2 IFN-γ ELISpot assay 1 and 6 months after the second vaccination.

2.3 *GNB3* c.825C>T genotyping

Genomic DNA was extracted from 200 µl EDTA blood using the QIAamp® DNA Blood Mini Kit (Qiagen, Hilden, Germany). Polymerase chain reaction (PCR) was performed with 2 µl genomic DNA and 30 µl Taq DNA-Polymerase 2x Master Mix Red (Ampliqon, Odense, Denmark) under the following conditions: initial denaturation 94°C for 3 min, 38 cycles with denaturation at 94°C for 30 s, annealing at 60°C for 30 s, elongation at 72°C for 30 s each, and final elongation at 72°C for 10 min (forward primer: 5′ GCCCTCAGTTCTTCC CCAAT 3'; reverse primer 3′ CCCACACGCTCAGACTTCAT 5′). PCR products were digested with BseDI (Thermo Scientific, Dreireich, Germany), and restriction fragments were analyzed by agarose gel electrophoresis. For the various genotypes, results from restriction fragment length polymorphism (RFLP)-PCR were validated by Sanger sequencing.

2.4 Detection of antibodies against SARS-CoV-2 spike protein

Determination of anti-spike SARS-CoV-2 antibody concentrations was performed using the SARS-CoV-2 S1 receptor-binding domain (RBD) IgG/sCOVG test (Siemens Healthcare GmbH, Erlangen, Germany) according to the manufacturer's instructions. Anti-Spike SARS-CoV-2 antibody concentration results were reported in binding antibody units per ml (BAU/ml). The limit of detection for positivity was 21.8 BAU/ml.

2.5 Detection of antibodies against the SARS-CoV-2 nucleocapsid protein

All samples were also analyzed for SARS-CoV-2 IgG antibodies against the nucleocapsid protein to exclude participants with prior SARS-CoV-2 infection. The Architect i2000SR CoV-2 IgG assay (Abbott Diagnostics, IL, United States) was used according to the manufacturer's instructions. Results with an index ≥1.4 were considered evidence of the previous infection.

2.6 ELISpot assay

To assess SARS-CoV-2-specific cellular immunity, we performed ELISpot assays using an overlapping peptide pool of SARS-CoV-2 S1 proteins (Miltenyi Biotec, Bergisch Gladbach, Germany) without the addition of any cytokines. We tested 250,000 peripheral blood mononuclear cells (PBMCs) per sample and measured IFN-γ production after 20 h of incubation. Mean values of duplicate cell cultures were considered. The median and mean spot numbers of autologous (unstimulated) controls were 0 and 0.05, respectively. SARS-CoV-2-specific spots were determined as stimulated minus unstimulated values (spots increment). The cut-off definition for positive results was based on negative control values (non-stimulated cultures) and the consideration that three times higher values for stimulated versus non-stimulated cells in cellular assays are often interpreted as a positive T-cell response. Using these criteria, the cut-off was 1.5 spot increments. Further details on the ELISpot assay and the cut-off definition have been published previously (Schwarzkopf et al., 2021).

2.7 Statistical analysis

Statistical analysis was performed with GraphPad Prism 7 (Graph Pad Software, San Diego, California, United States) and IBM SPSS Statistics 27 (IBM Software, Ehningen, Germany). Comparisons between three groups were made using the Kruskal–Wallis test and between two groups using the Mann–Whitney test. For genetic associations, we calculated the odds ratio (OR) and 95% confidence interval (CI) by Fisher's exact test using the Baptista–Pike method for the OR. *p*-values are given two-sided and values <0.05 were considered significant.

3 Results

3.1 Descriptive statistics of study groups

In the main study group of 204 participants, 105 participants were tested 1 month and 163 participants 6 months after the second vaccination with mRNA-1273. At both time points, sixty-four subjects participated. At month 1, the median age of the study group was 24 years (range 18–39), the BMI 22.5 kg/m^2 (range 17.0–29.9), and 69.5% (n=73) of participants were female. At month 6, the median age was 26 years (range 18–40), the BMI was 22.5 kg/m^2 (range 17.0–29.1), and 75.5% (n=123) were female. In the additional study group of 597 participants, the median age was 28 years (range 18–40), the BMI was 23.0 kg/m^2 (range 16.7–53.8), and 74.4% (n=444) of the participants were female.

3.2 Antibody titer against SARS-CoV-2 S1-RBD and T-cell response to SARS-CoV-2 S1 ELISpot assay one and six months after the second vaccination with mRNA-1273

The median antibody titer against SARS-CoV-2 S1-RBD was 3,887 BAU/ml (range 1,058–52,213) at month 1 (Figure 1A), which significantly ($p < 0.0001$) decreased to 644 BAU/ml (range 91–6,491) at month 6 (Figure 1B).

At month 1, 93.3% and at month 6, 41.7% of participants had a positive T-cell response in the SARS-CoV-2 IFN-γ ELISpot assay against the S1-peptide pool. Median values of spots increment decreased from 7.5 (range 0.5–60.5) to 0.5 (range -0.5–22.5) ($p < 0.0001$, Figures 1C,D).

Samples of 64 participants were available at both time points, 1 and 6 months after the second vaccination. The median antibody titer against SARS-CoV-2 S1-RBD was 3,682 BAU/ml (range 1,058–52,213) at month 1, which significantly ($p < 0.0001$) decreased to 731 BAU/ml (range 128–6,491) at month 6 (Figure 2A). Median values of spots increment in the SARS-CoV-2 IFN-γ ELISpot assay against the S1-peptide pool decreased from 6.0 (range 0.5–49) to 1.8 (range 0.0–22.5) ($p < 0.0001$, Figure 2B).

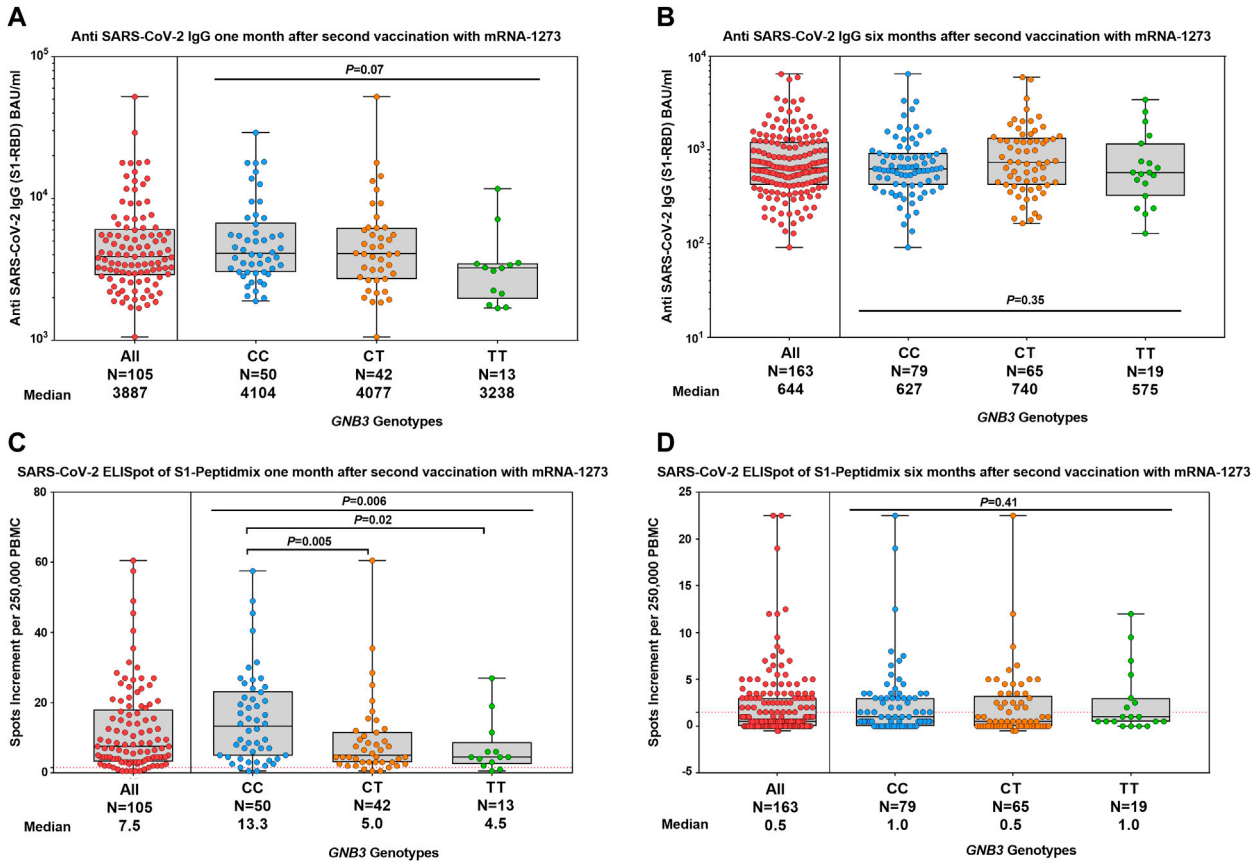


FIGURE 1
Humoral and cellular immune responses were stratified by *GNB3* genotypes 1 and 6 months after the second vaccination with mRNA-1273. Distribution of antibody concentrations against SARS-CoV-2 S1-RBD 1 month **(A)** and 6 months **(B)** after the second vaccination. ELISpot responses to the S1-protein of SARS-CoV-2 at 1 month **(C)** and 6 months **(D)** after the second vaccination. Antibody titers are reported in BAU/ml and T-cell response as spots increment. Red dashed lines indicate the cut-off for positivity (1.5 spots increment per 250,000 peripheral blood mononuclear cells).

3.3 Influence of *GNB3* c.825C>T on the immune response one and six months after the second vaccination with mRNA-1273

We investigated the impact of *GNB3* c.825C>T genotypes on the humoral immune response. We found a slightly albeit non-significantly lower anti-spike antibody titer in TT genotype carriers at month 1, which was no longer detectable in month 6 (Figures 1A,B). We validated these results in a larger cohort of 597 individuals (Figure 3).

At month 1, the median values of spots increment in the ELISpot assay were 13.3 (range 0.5–57.5) for CC, 5.0 (range 0.5–60.5) for CT, and 4.5 (range 0.5–27.0) for TT genotype carriers (p = 0.006, Figure 1C). CC genotype carriers had significantly increased T-cell responses compared to CT or TT genotypes (p = 0.005 and p = 0.02, respectively, Figure 1C). The effect was even more pronounced when comparing the CC genotype with T-allele carriers (13.3 vs. 4.5 spots increment, p = 0.001). At month 6, T-cell responses were strongly reduced and, therefore, genotype-dependent differences were no longer detectable (Figure 1D).

We analyzed the frequency distribution of *GNB3* genotypes above and below the median of 7.5 spots increment 1 month after the second vaccination to estimate if there is a genotype-related probability for a T-cell response above this cutoff. We found that CC genotype carriers had an almost 6-fold increased probability compared to TT genotype carriers (OR: 5.9, 95% CI: 1.6–21.5, p = 0.01) and an almost 3-fold increased probability compared to T-allele carriers (OR: 2.9, 95% CI: 1.3–6.2, p = 0.01) to mount a SARS-CoV-2-specific T-cell response above the median value.

4 Discussion

In this study, we observed a nearly 6-fold decrease in antibody titers from 1 to 6 months after the second

10.3389/fgene.2022.932043

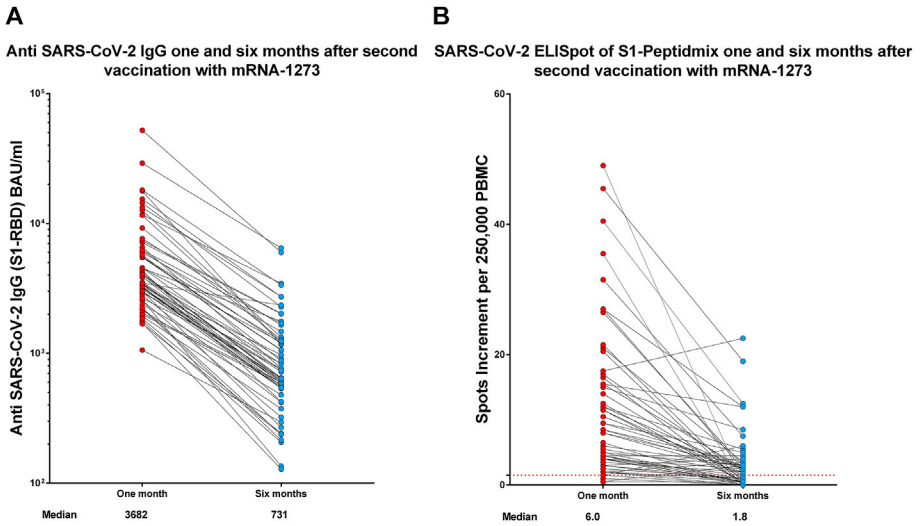


FIGURE 2
Humoral **(A)** and cellular **(B)** immune response 1 and 6 months after the second vaccination with mRNA-1273 in the group of 64 participants who were available at both time points. Antibody titers are reported in BAU/ml and T-cell response as spots increment. Red dashed lines indicate the cut-off for positivity (1.5 spots increment per 250,000 peripheral blood mononuclear cells).

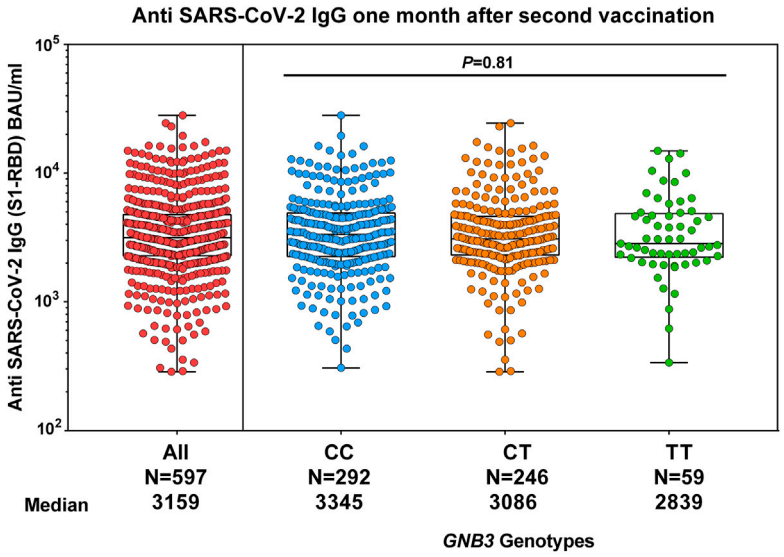


FIGURE 3
Comparison of humoral response and *GNB3* genotypes 1 month after the second vaccination with mRNA-1273 in the replication group. The median of anti-spike antibody levels is given in BAU/ml.

vaccination with mRNA-1273; nevertheless, all participants remained seropositive 6 months after the second vaccination. Despite many studies on the immune response after vaccination against COVID-19, there are only a few studies on the course of antibody titers over a longer period of time after vaccination with mRNA-1273. Those studies also reported a significant drop in antibody titers after vaccination with mRNA-1273 (Collier et al., 2021; Doria-Rose et al., 2021; Tre-Hardy et al., 2021; Gallagher et al., 2022).

In addition, Tré-Hardy *et al.* investigated whether different demographic characteristics such as age, BMI, or pre-existing conditions may influence the kinetics of antibody titers and

found no statistically significant relationship. Despite a very homogeneous study group of young, non-obese, and non-smoking participants without systemic immunosuppressive therapies or serious pre-existing conditions, we observed an almost 6-fold decrease in antibody titers, which confirms the findings of Tre-Hardy et al. (2021).

Some individuals generate lower antibody titers due to older age, pre-existing conditions, or immunosuppressive therapy (Geisen et al., 2021; Lindemann et al., 2021; Simon et al., 2021; Steensels et al., 2021; Widge et al., 2021). In our study, we present, for the first time, data on the potential influence of a G protein polymorphism on the immune response after vaccination with mRNA-1273. For this project, we chose the *GNB3* c.825C>T polymorphism because it was shown to correlate with T-cell responses to vaccination against HBV and to different recall antigens (Lindemann et al., 2001; Lindemann et al., 2002). In our current study, we observed that C-allele carriers had higher antibody titers, but this trend escaped statistical significance. In addition, no statistically significant differences were found between the genotypes of *GNB3* c.825C>T and the antibody titers after the booster vaccination against HBV (Lindemann et al., 2002). However, it has been shown that CT genotype carriers tend to have higher antibody titers after booster vaccination against HBV. It seems that *GNB3* c.825C > T may have a slight impact on the humoral immune response.

Data on T-cell kinetics after the vaccination with mRNA-1273 are scarce. Many studies tested T-cell immunity only once after vaccination or after a very short follow-up time. However, Gallagher *et al.* investigated the kinetics of T-cell responses after vaccination with mRNA-1273 at long-term follow-up and demonstrated approximately 30% decreased T-cell responses at a median of 223 days after the first vaccination with mRNA-1273 (Gallagher et al., 2022).

Our analysis of the cellular immunity also reveals a decrease in T-cell responses 6 months after the second vaccination with mRNA-1273. We observed a 15-fold decrease in T-cell responses in the SARS-CoV-2 ELISpot assay against the S1 peptide pool from 1 to 6 months after the second vaccination. In addition, at 6 months, only half of the participants had a T-cell response above the cut-off. It is also worth noting that our study is the first to measure T-cell responses in such a large cohort and, in addition, all tests were performed on freshly collected PBMC.

Our data show that CC genotype carriers have a stronger T-cell-mediated response and may be better protected against COVID-19 or have a milder COVID-19 infection after vaccination with mRNA-1273. This may also be an advantage for CC genotype carriers when antibodies cannot neutralize the virus and T-cell immunity is critical, e.g., after infection with immune escape variants of SARS-CoV-2 or when the humoral immune response is impaired.

However, at this time point, further studies are needed. First, our data should be replicated in an independent cohort. Further studies after booster vaccination causing a stronger immune response are also needed to see the influence of *GNB3* c.825C>T on the T-cell response after a longer follow-up. Last, the molecular mechanisms by which the *GNB3* c.825C>T polymorphism influences the T-cell response after SARS-CoV-2 vaccination and the potential clinical implications of these findings are to be yet unraveled.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

Ethics statement

The cohort study was reviewed and approved by the Ethics Committee of the Medical Faculty of the University of Duisburg-Essen (21–10005-BO). All participants provided their written informed consent to participate.

Author contributions

IC, BM, ML, and WS conceptualized and designed this study. BM obtained ethics approval. IC and BM obtained informed consent, demographic data, and samples from the participants, organized logistics, and performed laboratory analyses. Additionally, LT, NF, CE, and UD performed laboratory investigations. IC performed statistical analyses. IC, BM, ML, and WS drafted and wrote the manuscript, which was reviewed by all authors. All authors had full access to the data and approved submission of this manuscript. IC and BM had the final responsibility to submit this manuscript.

Funding

This work was supported by a grant (to IC) from the Stiftung Universitätsmedizin Essen of the Medical Faculty Essen. The funder of the study had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or in the decision of submitting the manuscript for publication. We acknowledge the support from the Open Access Publication Fund of the University of Duisburg-Essen.

Acknowledgments

We gratefully thank all participants of the study. We kindly thank Iris Manthey, Grit Müller, and Stephanie Büscher for technical support (Institute of Pharmacogenetics). We thank Lothar Volbracht and Marc Wichert for their support with the samples analyzed at the central laboratory (University Hospital Essen) and Frank Mosel (Department of Clinical Microbiology) for his support with coordination of the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.932043/full#supplementary-material

References

Collier, A. Y., Yu, J., McMahan, K., Liu, J., Chandrashekar, A., Maron, J. S., et al. (2021). Differential kinetics of immune responses elicited by covid-19 vaccines. *N. Engl. J. Med.* 385 (21), 2010–2012. doi:10.1056/NEJMc2115596

Crocchiolo, R., Gallina, A. M., Pani, A., Campisi, D., Cento, V., Sacchi, N., et al. (2022). Polymorphism of the HLA system and weak antibody response to BNT162b2 mRNA vaccine. *HLA* 99 (3), 183–191. doi:10.1111/tan.14546

Doria-Rose, N., Suthar, M. S., Makowski, M., O'Connell, S., McDermott, A. B., Flach, B., et al. (2021). Antibody persistence through 6 Months after the second dose of mRNA-1273 vaccine for covid-19. *N. Engl. J. Med.* 384 (23), 2259–2261. doi:10.1056/NEJMc2103916

Gallagher, K. M. E., Leick, M. B., Larson, R. C., Berger, T. R., Katsis, K., Yam, J. Y., et al. (2022). Differential T-cell immunity to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in mRNA-1273– and BNT162b2-vaccinated individuals. *Clin. Infect. Dis.*, ciac201. doi:10.1093/cid/ciac201

Geisen, U. M., Berner, D. K., Tran, F., Sumbul, M., Vullriede, L., Ciripoi, M., et al. (2021). Immunogenicity and safety of anti-SARS-CoV-2 mRNA vaccines in patients with chronic inflammatory conditions and immunosuppressive therapy in a monocentric cohort. *Ann. Rheum. Dis.* 80 (10), 1306–1311. doi:10.1136/annrheumdis-2021-220272

Gutierrez-Bautista, J. F., Sampedro, A., Gomez-Vicente, E., Rodriguez-Granger, J., Reguera, J. A., Cobo, F., et al. (2022). HLA class II polymorphism and humoral immunity induced by the SARS-CoV-2 mRNA-1273 vaccine. *Vaccines (Basel)* 10 (3), 402. doi:10.3390/vaccines10030402

Harrison, A. G., Lin, T., and Wang, P. (2020). Mechanisms of SARS-CoV-2 transmission and pathogenesis. *Trends Immunol.* 41 (12), 1100–1115. doi:10.1016/j.it.2020.10.004

Lindemann, M., Barsegian, V., Siffert, W., Ferencik, S., Roggendorf, M., and Grosse-Wilde, H. (2002). Role of G protein beta3 subunit C825T and HLA class II polymorphisms in the immune response after HBV vaccination. *Virology* 297 (2), 245–252. doi:10.1006/viro.2002.1467

Lindemann, M., Klisanin, V., Thummler, L., Fisenkci, N., Tsachakis-Muck, N., Ditschkowski, M., et al. (2021). Humoral and cellular vaccination responses against SARS-CoV-2 in hematopoietic stem cell transplant recipients. *Vaccines (Basel)* 9 (10), 1075. doi:10.3390/vaccines9101075

Lindemann, M., Virchow, S., Ramann, F., Barsegian, V., Kreuzfelder, E., Siffert, W., et al. (2001). The G protein beta3 subunit 825T allele is a genetic marker for enhanced T cell response. *FEBS Lett.* 495 (1-2), 82–86. doi:10.1016/s0014-5793(01)02339-0

Martinez-Flores, D., Zepeda-Cervantes, J., Cruz-Resendiz, A., Aguirre-Sampieri, S., Sampieri, A., and Vaca, L. (2021). SARS-CoV-2 vaccines based on the spike glycoprotein and implications of new viral variants. *Front. Immunol.* 12, 701501. doi:10.3389/fimmu.2021.701501

Schwarzkopf, S., Krawczyk, A., Knop, D., Klump, H., Heinold, A., Heinemann, F. M., et al. (2021). Cellular immunity in COVID-19 convalescents with PCR-confirmed infection but with undetectable SARS-CoV-2-specific IgG. *Emerg. Infect. Dis.* 27 (1), 122–129. doi:10.3201/2701.203772

Siffert, W., Rosskopf, D., Siffert, G., Busch, S., Moritz, A., Erbel, R., et al. (1998). Association of a human G-protein beta3 subunit variant with hypertension. *Nat. Genet.* 18 (1), 45–48. doi:10.1038/ng0198-45

Simon, D., Tascilar, K., Fagni, F., Kronke, G., Kleyer, A., Meder, C., et al. (2021). SARS-CoV-2 vaccination responses in untreated, conventionally treated and anticytokine-treated patients with immune-mediated inflammatory diseases. *Ann. Rheum. Dis.* 80 (10), 1312–1316. doi:10.1136/annrheumdis-2021-220461

Steensels, D., Pierlet, N., Penders, J., Mesotten, D., and Heylen, L. (2021). Comparison of SARS-CoV-2 antibody response following vaccination with BNT162b2 and mRNA-1273. *JAMA* 326 (15), 1533–1535. doi:10.1001/jama.2021.15125

Tre-Hardy, M., Cupaiolo, R., Wilmet, A., Antoine-Moussiaux, T., Della Vecchia, A., Horeanga, A., et al. (2021). Immunogenicity of mRNA-1273 COVID vaccine after 6 months surveillance in health care workers; a third dose is necessary. *J. Infect.* 83 (5), 559–564. doi:10.1016/j.jinf.2021.08.031

Widge, A. T., Rouphael, N. G., Jackson, L. A., Anderson, E. J., Roberts, P. C., Makhene, M., et al. (2021). Durability of responses after SARS-CoV-2 mRNA-1273 vaccination. *N. Engl. J. Med.* 384 (1), 80–82. doi:10.1056/NEJMc2032195

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 14 September 2022
DOI 10.3389/fgene.2022.975397

OPEN ACCESS

EDITED BY
Birte Möhlendick,
Essen University Hospital, Germany

REVIEWED BY
Bluma Brenner,
Lady Davis Institute (LDI), Canada
Marina Bobkova,
Ministry of Health Russian Federation, Russia

*CORRESPONDENCE
Penghui Shi,
shipenghui2021@163.com
Xinli Lu,
lxlii2009@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Genetics

RECEIVED 22 June 2022
ACCEPTED 23 August 2022
PUBLISHED 14 September 2022

CITATION
Fan W, Wang X, Zhang Y, Meng J, Su M, Yang X, Shi H, Shi P and Lu X (2022), Prevalence of resistance mutations associated with integrase inhibitors in therapy-naive HIV-positive patients in Baoding, Hebei province, China. *Front. Genet.* 13:975397. doi: 10.3389/fgene.2022.975397

Prevalence of resistance mutations associated with integrase inhibitors in therapy-naive HIV-positive patients in Baoding, Hebei province, China

Weiguang Fan[1†], Xiaodong Wang[2†], Yuchen Zhang[2†], Juan Meng[2], Miaomiao Su[2], Xuegang Yang[2], Haoxi Shi[1], Penghui Shi[1*] and Xinli Lu[3*]

[1]Clinical Laboratory, The People's Hospital of Baoding, Baoding, Hebei, China, [2]Infection Division, The People's Hospital of Baoding, Baoding, Hebei, China, [3]Department of AIDS Research, Hebei Provincial Center for Disease Control and Prevention, Shijiazhuang, Hebei, China

Antiretroviral therapy (ART) regimens containing integrase strand transfer inhibitors (INSTIs) are the recommended treatment for human immunodeficiency virus type 1 (HIV-1)-infected patients in the most recent guidelines in China. In this study, we investigated INSTI resistance mutations in newly diagnosed therapy-naive HIV-positive patients in Baoding City, Hebei Province (China) to provide guidance for implementing routine INSTI-associated HIV-1 genotypic resistance testing. Plasma samples were collected from HIV-1-infected patients without treatment at Baoding People's Hospital from January 2020 to December 2021. The part of HIV-1 *pol* gene encoding integrase was amplified, sequenced, and analyzed for INSTI resistance. Clinical data including demographic data, $CD4^+$ T cell counts, HIV-RNA loads, and resistance mutations were collected. Treatment-naïve HIV-1 patients (n = 131) were enrolled. We identified ten genotypes, and the predominant genotype was CRF01_AE in 67 patients (51.15%), CRF07_ BC in 39 patients (29.77%), subtype B in 11 patients (8.40%), and other subtypes (CRF68_01B, 3.82%; CRF55_01B, 1.53%, CRF80_0107, 1.53%; URFs 1.53%; and CRF103_01B, CRF59_01B, and CRF65_cpx, 1.4% each). Four major (E138A, R263k, G140S, and S147G) and three accessory (H51Y, Q146QL, and S153F) INSTI-resistance mutations were observed (genotype CRF01_AE, three patients; genotype B, one patient; and genotype CRF07_BC, one patient), resulting in different degrees of resistance to the following five INSTIs: raltegravir, elvitegravir, dolutegravir, bictegravir, and cabotegravir. The overall resistance rate was 3.82% (5/131). All INSTI-resistant strains were cross-resistant. The primary INSTI drug resistance rate among newly diagnosed HIV-infected patients in Baoding was low, but monitoring and research on HIV INSTI resistance should be strengthened in Baoding because INSTI-based regimen prescriptions are anticipated to increase in the near future.

KEYWORDS

HIV, drug resistance mutations, integrase strand transfer inhibitors, genotype, baoding city

Introduction

Human immunodeficiency virus (HIV), the causative agent of acquired immune deficiency syndrome (AIDS), was first reported in 1981, and it quickly became a major epidemic threatening human health (Centers for Disease Control and Prevention, 1981). Today, antiretroviral therapy (ART) has transformed AIDS from a fatal disease into a treatable but currently incurable chronic disease, and there have been significant changes in the life expectancy of people living with HIV (Trickey et al., 2017). However, with the widespread application of antiretroviral drugs in clinical practice, HIV drug resistance has become an important factor affecting ART efficacy (Godfrey et al., 2017; Zhao et al., 2017). The HIV resistance rate to non-nucleoside reverse transcriptase inhibitors has increased significantly, and some areas have even experienced a high rate of transmissible resistance, leading to incomplete viral suppression and treatment failure (Wittkop et al., 2011; Clutter et al., 2016; World Health Organization, 2019). One of the newest antiretroviral drug classes is integrase strand transfer inhibitors (INSTIs), which target the HIV integrase enzyme and block incorporation of reverse transcribed proviral DNA into the host genome. Because they have excellent tolerability, minimal toxicity, high efficacy, and are easy to use, integrase inhibitors became preferred agents for treatment-naive or experienced patients, and they are a novel treatment option for acquired and transmitted resistance in combination with other HIV drug classes (Winans and Goff, 2020; Boyd and Donovan, 2013; NAMSAL ANRS 12313 Study Gro up Kouanfack et al., 2019; Durham and Chahine, 2021). The United States Food and Drug Administration (United States FDA) has approved the following five drugs for clinical use: dolutegravir (DTG), raltegravir (RAL), elvitegravir (EVG), bictegravir (BIC), and cabotegravir (CAB). Despite integrase inhibitors plays an effective role in antiretroviral action with a novel mechanism of action, resistance is inevitable (Jiang et al., 2016), and previously published studies have shown that there were important INSTIs resistance mutations in newly diagnosed HIV-1 patients in Spain, Canada, and the United States (Stekler et al., 2015; Ji et al., 2018; Casadellà et al., 2020; López et al., 2021). Major INSTIs resistance mutations have also been discovered in Yunnan (Deng et al., 2019) and Jiangsu (Yin et al., 2021) in China, and INSTIs resistance mutations were found in our neighboring provinces such as Henan (Yang et al., 2021), Beijing (Yu et al., 2022). Therefore, early monitoring of INSTIs drug resistance is of great significance for clinical development of HIV medication guidance and timely adjustment of medication regimens. However, there have been no previous reports on the spread of INSTIs drug resistance strains in Hebei. Therefore, this study was a preliminary analysis of primary integrase gene mutation and drug resistance in Baoding to provide a reference for preventing and treating HIV-1 patients in China.

Materials and methods

Study participants

Baoding People's Hospital is the designated hospital of ART and takes charge of anti-HIV therapy of all HIV-1-infected individuals in Baoding city. In this study, a total of 131 HIV-1-infected individuals were recruited in Baoding City before starting ART from January 2020 to December 2021. Their blood samples were collected in Baoding People's Hospital, and written informed consent was obtained from all subjects before blood collection. The baseline demographic characteristics were investigated using face-to-face interviews when we collected subjects' blood samples. $CD4^+$ T cell counts were determined using the BD FACSCount system (Becton Dickenson, CA, United States). Plasma HIV RNA levels were quantitatively tested using the Ampliform HIV-1 Monitor Test, version 1.5 (Roche, Cobas AmpliPrep/TaqMan 48, Switzerland). The detection limit threshold was <20 copies/ml.

Nucleotide acid purification and polymerase chain reaction amplification

Viral RNA was purified using the QIAamp Viral RNA Mini Accessory Set (Qiagen, Duesseldorf, Germany), in accordance with the manufacturer's instructions. The part of HIV-1 *pol* gene encoding integrase (HXB2: 4053-5214) was amplified in two steps using self-designed primers. For the first-round of reverse transcriptase-polymerase chain reaction (RT-PCR) amplification, we used the M-MLV 4 One-Step RT-PCR kit (Beijing Bomaide Gene Technology, Beijing, China), in accordance with the manufacturer's instructions and using the following primers: INF12-2:5′-GCATTAGGRATYATTCARGCAC-3′ (outer l forward), INF12-1: 5′-GGRATYATTCARGCACAACCAG-3′ (outer forward), and INR15-1: 5′-TGGGATRTGTACTTCYGARCTTA-3′ (outer reversal). Thermal cycling conditions for the first round of RT-PCR consisted of reverse transcription at 50°C for 45 min, inactivation at 95°C for 2 min, which was followed by three cycles

of amplification at 95°C for 20 s, 50°C for 30 s, and 72°C for 2 min and 30 s. This was followed by 35 amplification cycles 95°C for 15 s, 50°C for 20 s, and 72°C for 2 min, with a final extension at 72°C for 10 min. Using the first-round RT-PCR product as a template, the second-round of PCR was performed using TaKaRa Premix Taq (TaKaRa Biotechnology, Dalian, China). Amplification was performed using the following primers: INF09:5′-TCTAYCTGKCATGGGTRCCAGCAC-3′ (inner forward) and INR18:5′-CATCCTGTCTACYTGCCACAC-3′ (inner reversal). The PCR conditions were as follows: a denaturing step at 95°C for 4 min, followed by three cycles at 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min and 30s; then 35 amplification cycles at 95°C for 20 s, 55°C for 20 s, and at 72°C for 2 min; and a final extension at 72°C for 10 min. Positive products were detected using 1.2% agarose gel electrophoresis, and the positive products with the same size, as shown by the electrophoresis bands, were purified and sequenced using the Sanger sequencing technology by the Beijing Bomaide Gene Technology Co., Ltd. (Beijing, China). The sequencing primers were selected from the second round of amplification primers INF09, INR18 and KVL082: 5′-GGVATTCCCTATCAATCCCCAAAG-3′, KVL083: 5′-GAATACTGCCATTTGTACTGCTG-3′.

Sequence and drug-resistant mutation analysis

Raw sequences were assembled using Contig Express 9.1. Multiple sequence alignment was completed using ClustalW, and manual editing was performed using Bio-Edit 7.0 software. A neighbor-joining phylogenetic tree was constructed using the Kimura two-parameter model with 1000 bootstrap replicates using Molecular Evolutionary Genetic Analysis version 6.0 software (MEGA 6.0). HIV-1 subtypes were preliminarily analyzed using the online REGA HIV-1 Subtyping Tool 3.0 (http://dbpartners.stanford.edu:8080/RegaSubtyping/stanford-hiv/typingtool/) and identified using a neighbor-joining tree based on the HIV-1 integrase gene sequences. The Stanford HIV-1 drug resistance database (HIVdb version 9.0) was used to analyze the mutations. The Stanford HIVdb algorithm classifies the likelihood of INSTI resistance mutations being associated with resistance into “susceptible”, “potential low-level”, “low-level”, “intermediate”, and “high-level.” In our study, sequences containing low-level to high-level gene mutations were defined as resistance sequences. The related sequences were submitted to GenBank, and the accession numbers are ON529336 - ON529466.

Statistical analysis

Statistical analysis was conducted using SPSS 21.0 (IBM Corp., Armonk, NY, United States). Means or frequencies were used to summarize demographic data. The differences of polymorphisms mutations according to subtypes were analyzed using chi-square test. All tests were two-tailed, and p-values <0.05 were considered statistically significant.

TABLE 1 Clinical characteristics of patients at enrolment.

Variables	Patients N (%)
Age (years)	
18–50	113 (86.26)
51–76	18 (13.74)
Sex	
Male	115 (87.79)
Female	16 (12.21)
Transmission route	
MSM	115 (87.79)
Heterosexual	16 (12.21)
CD^4T-cell count (cells/μL)	
<200	45 (34.35)
≥200	86 (65.65)
HIV-1 viral load (log_{10} RNA copies/mL)	
<5	66 (50.38)
≥5	65 (49.62)
Genotype	
CRF01_AE	67 (51.15)
CRF07_BC	39 (29.77)
B	11 (8.40)
CRF68_01B	5 (3.82)
CRF55_01B	2 (1.53)
CRF80_0107	2 (1.53)
URFs	2 (1.53)
CRF103_01B	1 (0.76)
CRF65_cpx	1 (0.76)
CRF59_01B	1 (0.76)

MSM, men who have sex with men; HIV, human immunodeficiency virus.

Results

Patient characteristics

One hundred thirty-one patients were evaluated in this study. Most were men (115/131, 87.79%), and the median age was 34 years (range, 18–76 years). The most common transmission routes were among men who have sex with men (MSM) (87.79%, 115/131) followed by heterosexual (HET) transmission (12.21%, 16/131). The median $CD4^+$ T cell count was 166 cells/μl (range, 4–926 cells/μl). HIV-1 RNA loads ranged from 4.72 log10 RNA copies to 7.57 log10 RNA copies/ml (Table 1). Phylogenetic analysis of the integrase (IN) gene revealed that the sequenced strains could be divided into ten genotypes (Figure 1). The prevalence of each subtype was as follows: CRF01_AE

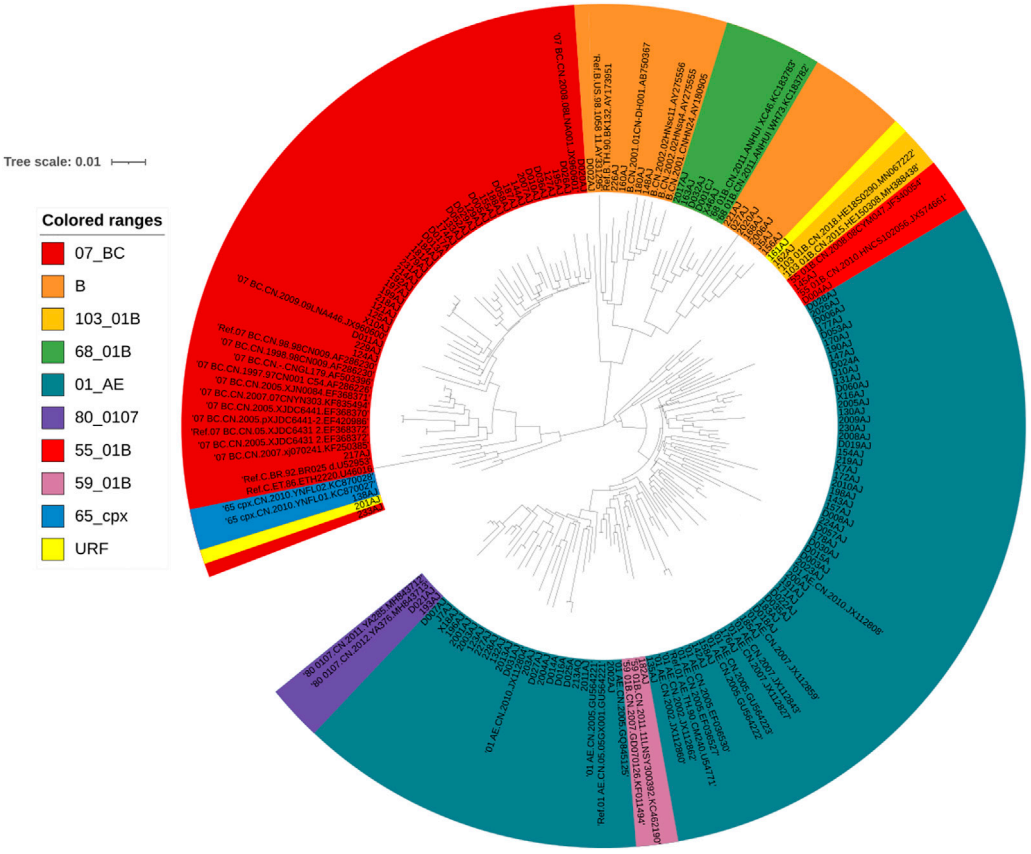


FIGURE 1
Phylogenetic tree analysis based on the HIV-1 integrase gene sequences. A neighbor-joining tree was constructed using MEGA 6.0 with 1000 bootstrap replicates, and were adjusted using the online itol (https://itol.embl.de/). The standard reference sequences of HIV-1 subtypes were downloaded from the HIV database (http://www.hiv.lanl.gov/content/index). Different subtypes are shown in using different colors listed in this figure.

(51.15%, 67/131), CRF07_BC (29.77%, 39/131), B (8.40%,11/131), and other subtypes (10.69%, 14/131).

INSTI resistance mutations

Drug resistance analyses showed that four major INSTI-resistance mutations (E138A, G140S, S147G, and R263K) and three accessory INSTI-resistance mutations (H51Y, Q146QL, and S153F) were detected in five patients. Information about INSTI resistance mutations corresponding to the specific drugs are listed in Table 2. To date, the CAB-associated resistance mutations have been listed in the HIVdb version 9.0. On the basis of this classification system, 96.18% of participants had viruses that were fully susceptible to five INSTIs. Low-level resistance to second-generation INSTIs DTG, BIC, and CAB were identified in 1.53% (2/131), 1.53% (2/131), and 2.29% (3/131) of participants, respectively. Intermediate-level resistance to all three second-generation INSTIs was 0.76% (1/131), and there was no high-level resistance. However, for both the first-generation INSTIs EVG, and RAL, low-grade resistance was 2.29% (3/131) and intermediate-level resistance was 0.76% (1/131). There was one case of high resistance to EVG. The overall INSTIs drug resistance rate was 3.82% (5/131) in this study.

Natural polymorphisms of CRF01_AE and CRF07_BC strains at integrase resistance-related sites

Compared with the international standard strain HXB2 of subtype B, the CRF01_AE and CRF07_BC strains had more naturally occurring polymorphic variants in the encoded amino acids in the integrase gene coding region. The CRF01_AE strain had more than 50% mutation rates at K14R, V31I, I72V, T112V, T124A, T125A, G134N, I135V, K136Q, D167E, V201I, L234I/S,

TABLE 2 Characteristics of five ART-naive individuals with INSTI drug resistance mutations.

Patients CodeID	Age (years)	Sex	Transmission Route	Genotype	CD4 count (cells/?l)	Viral Load (log10 RNA copies/ml)	Integrase Mutations		Drug resistance				
							Major	accessory	BIC	DTG	CAB	EVG	RAL
130AJ	17	male	MSM	01_AE	416	5.34		H51Y	P	P	L	L	L
148AJ	56	male	MSM	B	542	4.71	E138A		P	P	P	L	L
178AJ	50	male	MSM	01_AE	362	5.58		S153F	L	L	L	L	P
192AJ	19	male	MSM	07_BC	868	4.37	R263K		I	I	I	I	L
2004AJ	26	male	MSM	01_AE	8	4.75	G140S,S147G	Q146QL	L	L	L	H	I

ART, anti-retroviral therapy; INSTI, integrase strand-transfer inhibitor; MSM, men who have sex with men; BIC, bictegravir; DTG, dolutegravir; CAB, cabotegravir; EVG, elvitegravir; RAL, raltegravir; P, potential low-level resistance; L, low-level resistance; I, intermediate resistance; H, high-level resistance.

and S283G, while the CRF07_BC strain had K42Q, L101I, T112V, T124A, T125A, I135V, K136Q, V201I, R269K, D278A, and S283G have a mutation rate of more than 50%. There were statistically significant differences between the two genotype mutation rates at 14 polymorphic variants K14R, K42Q, I72V, L101I, T125A, G134N, I135V, K136Q, D167E, V201I, L234I, L234S, R269K, and D278A ($P < 0.05$; Table 3).

Discussion

Since RAL (the first INSTI) was approved by the US FDA in 2007, HIV treatment has entered a new era. ART regimens containing INSTIs have become the recommended treatment for HIV-1-infected patients in the Chinese Guidelines for Diagnosis and Treatment of HIV/AIDS (2021) in China (AIDS and Hepatitis C Professional Group, 2021). However, with the widespread use of INSTIs, INSTI-related drug resistance mutations have gradually emerged. This study monitored the prevalence of resistance mutations associated with HIV INSTI in ART-naïve patients in Baoding from January 2020 to December 2021. Four major INSTI-resistant mutations were detected in five patients, including E138A, G140S, S147G, and R263K as well as three accessory INSTI-resistant mutations H51Y, Q146QL, and S153F, which cause varying degrees of resistance in the following five INSTIs: BIC, DTG, CAB, EVG, and RAL (Table 2). The prevalence of INSTI resistance was 3.82% (5/131). This INSTI resistance prevalence was higher than that in Jiangsu (0.76%) (Yin et al., 2021), Henan (1.7%) (Yang et al., 2021), and Beijing (0.34%) in China (Yu et al., 2022) as well as in Italy (0.2%) (Rossetti et al., 2018), Austria (2.3%) (Zoufaly et al., 2017), and Uganda (1.2%) (McCluskey et al., 2021), but lower than that in Yunnan (5.7%) (Deng et al., 2019) and Poland (8.3%) (Parczewski et al., 2012) and close to that in Korea (3.4%) (Jeong et al., 2019), which suggests that it is necessary for us to strength the surveillance of INSTIs resistance in order to control and prevent the spread of HIV-1 resistant strains.

In this study, all INSTI-resistant strains were resistant to EVG, and one patient was highly resistant to EVG. In the subtype distribution, there were three patients with the CRF01_AE genotype, which is carried in the H51Y, S153SF, G140S, S147SG, and Q146QL mutations, and it had high, intermediate, and low levels of resistance to five INSTIs. One patient who had subtype B carried E138A and had low resistance to RAL and EVG, and one patient with the CRF07_BC subtype carried the R263K mutation and had intermediate resistance to BIC, DTG, CAB, and EVG but low resistance to RAL. R263K is selected *in vitro* using EVG, DTG, and BIC and in patients receiving DTG. It reduces DTG and BIC susceptibility approximately two-fold and EVG susceptibility to a greater extent (https://hivdb.

TABLE 3 Natural polymorphisms of CRF01_AE and CRF07_BC strains at integrase resistance-related sites.

polymorphic variants	CRF01_AE		CRF07_BC		P	polymorphic variants	CRF01_AE		CRF07_BC		*P*
	N	%	N	%			N	%	N	%	
K14R	45	67.16	10	25.64	<0.001	I135V	60	89.55	28	71.79	0.019
V31I	42	62.69	18	46.15	0.098	K136Q	62	92.54	30	76.92	0.022
K42Q	5	7.46	28	71.79	<0.001	D167E	42	62.69	5	12.82	<0.001
I72V	34	50.75	6	15.38	<0.001	V201I	67	100	33	84.62	0.001
L101I	18	26.87	30	76.92	<0.001	L234I	62	92.54	7	17.95	<0.001
T112V	63	94.03	34	87.18	0.222	L234S	2	2.99	24	61.54	<0.001
T124A	64	95.52	34	87.18	0.117	R269K	5	7.46	24	61.54	<0.001
T125A	65	97.01	31	79.49	0.003	D278A	6	8.96	27	69.23	<0.001
G134N	61	91.04	8	20.51	<0.001	S283G	61	91.04	32	82.05	0.173

Note: The differences of polymorphisms mutations according to subtypes were analyzed using chi-square test.

stanford.edu/dr-summary/comments/INSTI.[OL]/). E138A is non-polymorphic mutation that is selected-for in patients receiving RAL, EVG, and DTG. Alone, this non-polymorphic mutation does not reduce INSTI susceptibility, but when it occurs in combination with the Q148 mutations, it is associated with high-level resistance to RAL and EVG and an intermediate reduction to DTG and BIC susceptibility (https://hivdb.stanford.edu/dr-summary/comments/INSTI.[OL]/). However, this study showed that none of these mutations occurred in the E138A and Q148 mutation combinations. In the past years, INSTIs have never been included in the free ART regimens in Hebei. We infer that HIV-1 INSTI resistant strans should be introduced into Hebei from other areas.

In this study, many amino acid polymorphisms that are different from the standard B subtype strains were detected in both the CRF01_AE and CRF07_BC strains, which have also been reported in untreated persons who are infected with other genotype strains in other countries (Brado et al., 2018; Alaoui et al., 2019; Huang et al., 2019; Lai et al., 2021). Although these polymorphic variants were encountered in more than half of the studied participants, other polymorphic mutations that probably have no effect on INSTIs susceptibility were reported in previous *in vitro* studies, and clinical trials showed different frequencies (Kobayashi et al., 2011; Casadellà et al., 2015). However, naturally occurring polymorphisms impact the intasome complex stability and may, therefore, contribute to the overall potency against INSTIs and natural polymorphisms, while subtype-specific differences may influence the effect of individual treatment regimens (Brado et al., 2018). The differences of polymorphic variants between CRF01_AE and CRF07_BC suggest that the disease progress of CRF01_AE is faster than CRF07_BC (Ye et al., 2022).

There are some limitations to our study, such as using only the IN gene in the genotyping accuracy and CRF identification. The lacking of reverse transcriptase and protease resistance data not only affects genotyping accuracy but also the assessment of resistance to first-line regimens. Additionally, the relatively small number of samples obtained for our analysis limits our ability to definitively provide frequency analysis for INSTI resistance mutations. Study results are also limited because Sanger sequencing was used instead of next-generation sequencing, which allows minority variants to be reported. In our future study, we will increase the sample size and sequence the whole length *pol* gene using more sensitive techniques such as next-generation sequencing methods to minimize the limitations.

Conclusion

The respective prevalence rate of INSTI major resistance mutations in ART-naive patients in Baoding was 3.82%. We think that the surveillance of INSTI resistance should be recommended before regarding treatment regimens containing INSTI are planned in Baoding city. The study is small and a more clear picture of INSTI resistance in China should be evaluated in a larger national study.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee at the Peoples Hospital of Baoding 2019-03. Written informed consent to participate in this study was provided by the patient/participants or patient/participants legal guardian/next of kin.

Author contributions

WF, XW, and XL designed the study. WF and PS acquired the sequences. JM, MS, XY, and HS analyzed and interpreted the data. WF, XW, YZ, and XL wrote the manuscript. All authors read and agreed to the published version of the manuscript.

Funding

This study was supported by the science and technology planning project of Baoding prefecture (1951ZF007).

Acknowledgments

The authors thank all staff involved in this study. We thank Liwen Bianji (Edanz) (www.liwenbianji.cn) for editing the English text of a draft of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

AIDS and Hepatitis C Professional Group. (2021) The Chinese guidelines for diagnosis and treatment of HIV/AIDS (2021)[J]. *Chin. J. AIDS STD* 27(11), 1182–1201.

Alaoui, N., El Alaoui, M. A., El Annaz, H., Farissi, F. Z., Alaoui, A. S., El Fahime, E., et al. (2019). HIV-1 integrase resistance among highly antiretroviral experienced patients from Morocco. *Intervirology* 62 (2), 65–71. doi:10.1159/000501016

Boyd, M. A., and Donovan, Donovan B. (2013). Antiretroviral therapy: Dolutegravir sets SAIL(ING). *Lancet (London, Engl.* 382, 664–666. doi:10.1016/S0140-6736(13)61456-7

Brado, D., Obasa, A. E., Ikomey, G. M., Cloete, R., Singh, K., Engelbrecht, S., et al. (2018). Analyses of HIV-1 integrase sequences prior to South African national HIV-treatment program and available of integrase inhibitors in Cape Town, South Africa.Erratum in. *Sci. Rep.Sci Rep.* 88 (11), 47096262. doi:10.1038/s41598-018-22914-5

Casadellà, M., Santos, J. R., Noguera-Julian, M., Mican-RiveRa, R., Domingo, P., AntelA, A., et al. (2020). Primary resistance to integrase strand transfer inhibitors in Spain using ultrasensitive HIV-1 genotyping. *J. Antimicrob. Chemother.* 75 (12), 3517–3524. doi:10.1093/jac/dkaa349

Casadellà, M., van Ham, P. M., Noguera-Julian, M., vAn Kessel, A., Pou, C., Hofstra, L. M., et al. (2015). Primary resistance to integrase strand-transfer inhibitors in Europe. *J. Antimicrob. Chemother.* 70 (10), 2885–2888. doi:10.1093/jac/dkv202

Centers for Disease Control and Prevention (1981). Pneumocystic pneumonia-los angeles [J]. *MMWR Morb. Mortal. Wkly. Rep.* (3021), 1–3.

Clutter, D. S., Jordan, M. R., Bertagnolio, S., and Shafer, R. W. (2016). HIV-1 drug resistance and resistance testing. *Infect. Genet. Evol.* 46, 292–307.

Deng, X. M., Liu, J. F., and Zhang, M. (2019). Mutations of primary integrase gene resistance of HIV/AIDS patients in Yunnan province. *Clin. J. AIDS STD* 25 (4), 327–341.

Durham, S. H., and Chahine, E. B. (2021). Cabotegravir-Rilpivirine: The first complete long-acting injectable regimen for the treatment of HIV-1 infection. *Ann. Pharmacother.* 55 (11), 1397–1409. doi:10.1177/1060028021995586

Godfrey, C., Bobkova, M., Boucher, C., Ravasi, G., Chen, P., Zhang, F., et al. (2017). Regional challenges in the prevention of human immunodeficiency virus drug resistance. *J. Infect. Dis.* 216 (9), S816–S819. doi:10.1093/infdis/jix408

Huang, X. T., Sun, Z. S., An, M. H., Zhao, L., Wang, L., Ding, H. B., et al. (2019). Primary drug resistance to integrase inhibitors among newly diagnosed HIV infected patients in Shenyang city. *Chin. J. Clin. Lab. Sci.* 10, 721–725.

Jeong, W., Jung, I. Y., Choi, H., Kim, J. H., Seong, H., Ahn, J. Y., et al. (2019). Integrase strand transfer inhibitor resistance mutations in antiretroviral therapy-naive and treatment-experienced HIV patients in South Korea. *AIDS Res. Hum. Retroviruses* 35, 213–216. doi:10.1089/AID.2018.0213

Ji, H., Patterson, A., Taylor, T., Rank, C., Halverson, J., Capina, R., et al. (2018). Prevalence of primary drug resistance against HIV-1 integrase inhibitors in Canada. *J. Acquir. Immune Defic. Syndr.* 78 (1), e1–e3. doi:10.1097/QAI.0000000000001649

Jiang, J., Xu, X., Guo, W., Su, J., Huang, J., Liang, B., et al. (2016). Dolutegravir(DTG, S/GSK1349572) combined with other ARTs is superior to RAL- or EFV-based regimens for treatment of HIV-1 infection: A meta-analysis of randomized controlled trials. *AIDS Res. Ther.* 13 (1), 30. doi:10.1186/s12981-016-0115-x

Kobayashi, M., Yoshinaga, T., Seki, T., Wakasa-Morimoto, C., Brown, K. W., Ferris, R., et al. (2011). *In vitro* antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* 55 (2), 813–821. doi:10.1128/AAC.01209-10

NAMSAL ANRS 12313 Study Group, Kouanfack, C., Mpoudi-Etame, M., Omgba Bassega, P., Eymard-Duvernay, S., and Leroy, S. (2019). Dolutegravir-based or low-dose efavirenz-based regimen for the treatment of HIV-1. *N. Engl. J. Med. Overseas. Ed.* 381, 816–826. doi:10.1056/nejmoa1904340

Lai, J., Liu, Y., Han, X., Huang, A., Lin, J., Ao, W., et al. (2021). Low frequency of integrase inhibitor resistance mutations among therapy-naïve HIV patients in southeast China. *Drug Des. devel. Ther.* 15, 889–894. doi:10.2147/DDDT.S286863

López, P., Tirado, G., Arias, A., Sanchez, R., Rodriguez-Lopez, E. R., and Rivera-Amill, V. (2021). Short communication: Integrase strand transfer inhibitors drug resistance mutations in Puerto Rico HIV-positive individuals. *Int. J. Environ. Res. Public Health* 18 (5), 2719. doi:10.3390/ijerph18052719

McCluskey, S. M., Kamelian, K., Musinguzi, N., Kigozi, S., Boum, Y., Bwana, M. B., et al. (2021). Pre-treatment integrase inhibitor resistance is uncommon in antiretroviral therapy-naive individuals with HIV-1 subtype A1 and D infections in Uganda. *AIDS* 35 (7), 1083–1089. doi:10.1097/qad.0000000000002854

Parczewski, M., Bander, D., Urbańska, A., and Boron-Kaczmarska, A. (2012). HIV-1 integrase resistance among antiretroviral treatment naive and experienced

patients from Northwestern Poland. *BMC Infect. Dis.* 12, 368. doi:10.1186/1471-2334-12-368

Rossetti, B., Di Giambenedetto, S., Torti, C., Postorino, M. C., Punzi, G., SaladiniF., et al. (2018). Evolution of transmitted HIV-1 drug resistance and viral subtypes circulation in Italy from 2006 to 2016. *HIV Med.* 19, 619–628. doi:10.1111/hiv.12640

Stekler, J. D., McKernan, J., Milne, R., Tapia, K. A., Mykhalchenko, K., Holte, S., et al. (2015). Lack of resistance to integrase inhibitors among antiretroviral-naive subjects with primary HIV-1 infection, 2007-2013. *Antivir. Ther.* 20 (1), 77–80. doi:10.3851/IMP2780

Trickey, A., May, M. T., Vehreschild, J-J, Obel, N., Gill, M. J., Crane, H. M., et al. (2017). Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: A collaborative analysis of cohort studies. *Lancet. HIV* 4, e349–e356. doi:10.1016/S2352-3018(17)30066-8

Winans, S., and Goff, S. P. (2020). Mutations altering acetylated residues in the CTD of HIV-1 integrase cause defects in proviral transcription at early times after integration of viral DNA. *PLoS Pathog.* 16, e1009147. doi:10.1371/journal.ppat.1009147

Wittkop, L., Günthard, H. F., de Wolf, F., Dunn, D., Cozzi-Lepri, A., de Luca, A., et al. (2011). Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): A European multicohort study. *Lancet. Infect. Dis.* 11 (5), 363–371. doi:10.1016/S1473-3099(11)70032-9

World Health Organization (2019). *HIV drug resistance report 2019*. Geneva: World Health Organization.

Yang, Z., Yang, X., Deng, X., Wei, S., Liu, J., Ma, J., et al. (2021). Prevalence of integrase strand transfer inhibitor (INSTIs) resistance mutations in Henan Province, China (2018-2020). *Infection* 49 (6), 1195–1202. doi:10.1007/s15010-021-01668-9

Ye, J., Chen, J., Wang, J., Wang, Y., Xing, H., Yu, F., et al. (2022). CRF07_BC is associated with slow HIV disease progression in Chinese patients. *Sci. Rep.* 12 (1), 3773. doi:10.1038/s41598-022-07518-4

Yin, Y. Q., Lu, J., Zhou, Y., Shi, L. E., Yuan, D. F., Chen, J. S., et al. (2021). Drug resistance to HIV-1 integrase inhibitors among treatment-naive patients in Jiangsu, China. *Biomed. Environ. Sci.* 34 (5), 400–403. doi:10.3967/bes2021.053

Yu, F., Li, Q., Wang, L., Zhao, H., Wu, H., Yang, S., et al. (2022). Drug resistance to HIV-1 integrase inhibitors among treatment-naive patients in beijing, China. *Pharmgenomics. Pers. Med.* 15, 195–203. doi:10.2147/PGPM.S345797

Zhao, Y., Zhang, M., Shi, C. X., Huang, J., Zhang, F., Rou, K., et al. (2017). Mortality and virological failure among HIV-infected people who inject drugs on antiretroviral treatment in China: An observational cohort study. *Drug Alcohol Depend.* 170, 189–197. doi:10.1016/j.drugalcdep.2016.11.011

Zoufaly, A., Kraft, C., Schmidbauer, C., and Puchhammer-Stoeckl, E. (2017). Prevalence of integrase inhibitor resistance mutations in Austrian patients recently diagnosed with HIV from 2008 to 2013. *Infection* 45, 165–170. doi:10.1007/s15010-016-0936-5

frontiers | Frontiers in Genetics

TYPE Case Report
PUBLISHED 02 February 2023
DOI 10.3389/fgene.2023.1105739

OPEN ACCESS

EDITED BY
Hezhao Ji,
Public Health Agency of Canada (PHAC), Canada

REVIEWED BY
Joan Fibla,
Universitat de Lleida, Spain
Binhua Liang,
Public Health Agency of Canada (PHAC), Canada

*CORRESPONDENCE
Weina An,
anweina2023@163.com
Xinli Lu,
lxlii2009@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Genetics

RECEIVED 23 November 2022
ACCEPTED 24 January 2023
PUBLISHED 02 February 2023

CITATION
Zhang B, Chen S, Meng J, Su M, Fan W, An W and Lu X (2023), Identification of two near-identical novel HIV-1 unique recombinant forms (CRF01_AE/B) among men who have sex with men in baoding, hebei, China.
Front. Genet. 14:1105739.
doi: 10.3389/fgene.2023.1105739

Identification of two near-identical novel HIV-1 unique recombinant forms (CRF01_AE/B) among men who have sex with men in baoding, hebei, China

Binbin Zhang[1†], Sisi Chen[2†], Juan Meng[3†], Miaomiao Su[3], Weiguang Fan[2], Weina An[2*] and Xinli Lu[4*]

[1]Laboratory of Endocrinology, Baoding No. 1 Central Hospital, Baoding, China, [2]Clinical Laboratory, The People's Hospital of Baoding, Baoding, Hebei, China, [3]Infection Division, The People's Hospital of Baoding, Baoding, Hebei, China, [4]Department of AIDS Research, Hebei Provincial Center for Disease Control and Prevention, Shijiazhuang, Hebei, China

Men who have sex with men (MSM) are the most frequent infection route of the human immunodeficiency virus (HIV) in Baoding, China, creating chances for the occurrence of unique recombinant forms (URFs) of the virus, i.e., recombination of different subtypes caused by co-circulation of multiple subtypes. In this report, two near-identical URFs (BDD002A and BDD069A) isolated from MSM in Baoding were identified. Phylogenetic tree analysis based on nearly full-length genomes (NFLGs) revealed that the two URFs formed a distinct monophyletic cluster with a bootstrap value of 100%. Recombinant breakpoints analysis identified that the NFLGs of BDD002A and BDD069A were both composed of CRF01_AE and subtype B, with six subtype B mosaic segments inserted into the CRF01_AE backbone. The CRF01_AE segments of the URFs clustered closely with the CRF01_AE reference sequences, and the B subregions clustered with the B reference sequences. The recombinant breakpoints of the two URFs were almost identical. These results suggest that effective interventions are urgently needed to prevent the formation of complex HIV-1 recombinant forms in Baoding, China.

KEYWORDS

HIV, near full-length genome, unique recombination forms, baoding, MSM

Introduction

The human immunodeficiency virus-1 (HIV-1) possesses an extremely high mutation frequency, resulting in HIV-1-enriched gene polymorphisms (Yebra et al., 2018). During the spread of HIV, if more than two subtypes infect the same cell, their genomic information can exchange to generate recombinant virus genomes (Moore and Hu, 2009). A total of 132 circulating recombinant forms (CRFs) of HIV and many unique recombinant forms (URFs) have been reported worldwide (https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/crfs). In recent years, CRF01_AE and subtype B have become the two main HIV-1 genotypes prevalent in key populations of China, especially among men who have sex with men (MSM) (He et al., 2012). Hebei is a northern province of China with low HIV prevalence (Lu et al., 2017). By the end of October 2020, 15,178 individuals in the Hebei Province were diagnosed with HIV-1/AIDS, and the number of individuals infected with HIV-1 through MSM reached 77.5% (Lu et al., 2020; Wang, 2020). Baoding, which borders Beijing and Tianjin, was the second region severely affected by HIV-1 in Hebei Province, with MSM

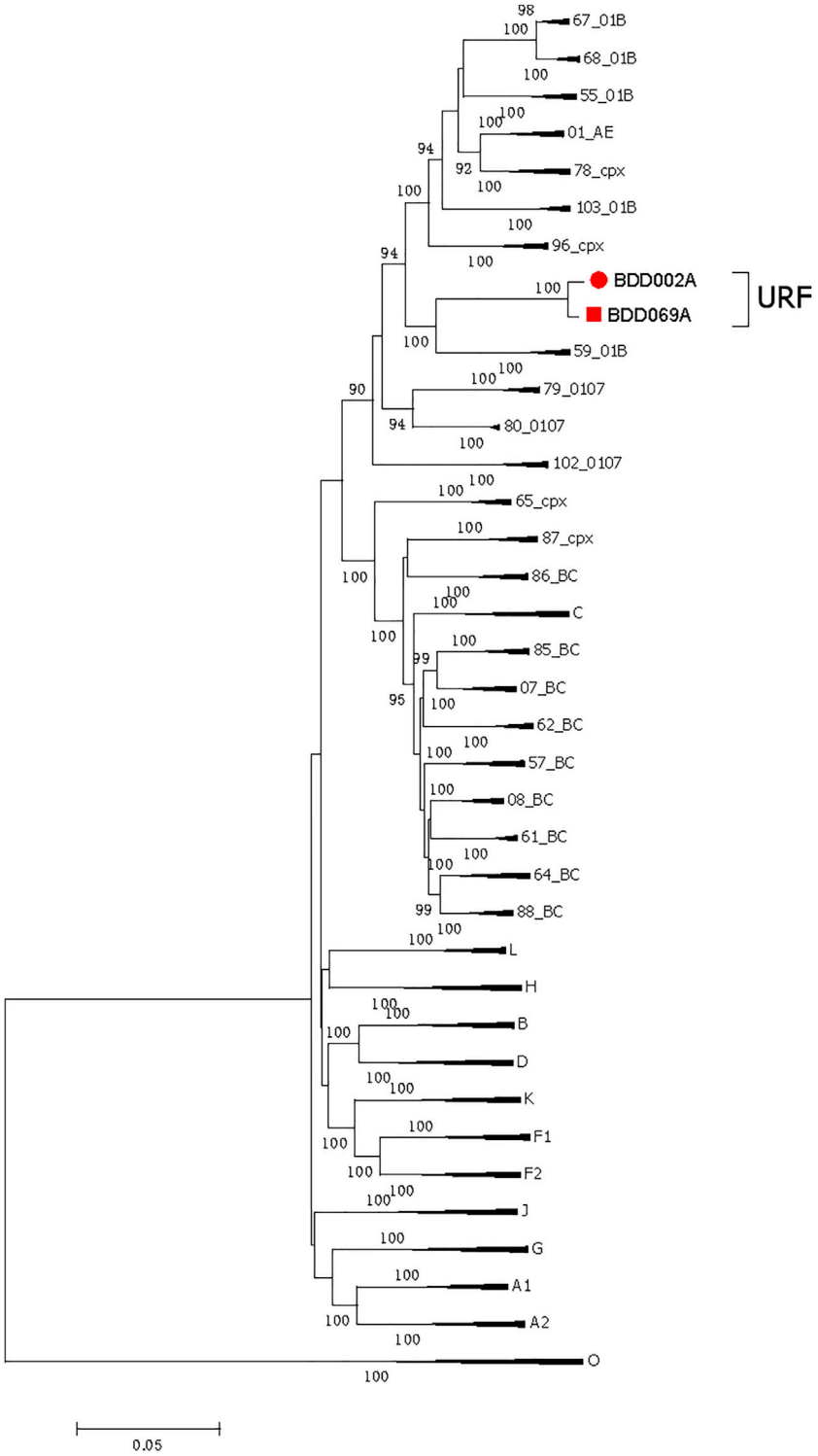


FIGURE 1
The phylogenetic tree is based on NFLG sequences BDD002A and BDD069A. The standard subtype references were downloaded from the Los Alamos National Laboratory HIV Database (www.hiv.lanl.gov). The neighbor-joining phylogenetic tree of BDD002A (8,810 bp, red-filled circle●) and BDD069A (8,944 bp, red-filled square■) was constructed based on the NFLG sequences using Mega6.0. The stability of each node was assessed by bootstrap tests with 1,000 replicates, and only bootstrap values ≥90% are shown at the corresponding nodes. The scale bar represents a 5% genetic distance. NFLG, near full-length genome.

responsible for 84.8% of the HIV-1-infected population (Shi et al., 2021a). The prevalence of subtypes CRF01AE and B was similar to that reported in the whole province at 49.44% and 17.78%, respectively (Shi et al., 2021b). In this study, we identified two highly similar, nearly full-length genome (NFLG) sequences isolated from MSM in Baoding, Hebei Province. These two unique recombinant forms (URFs) were composed of subtypes CRF01_AE and B.

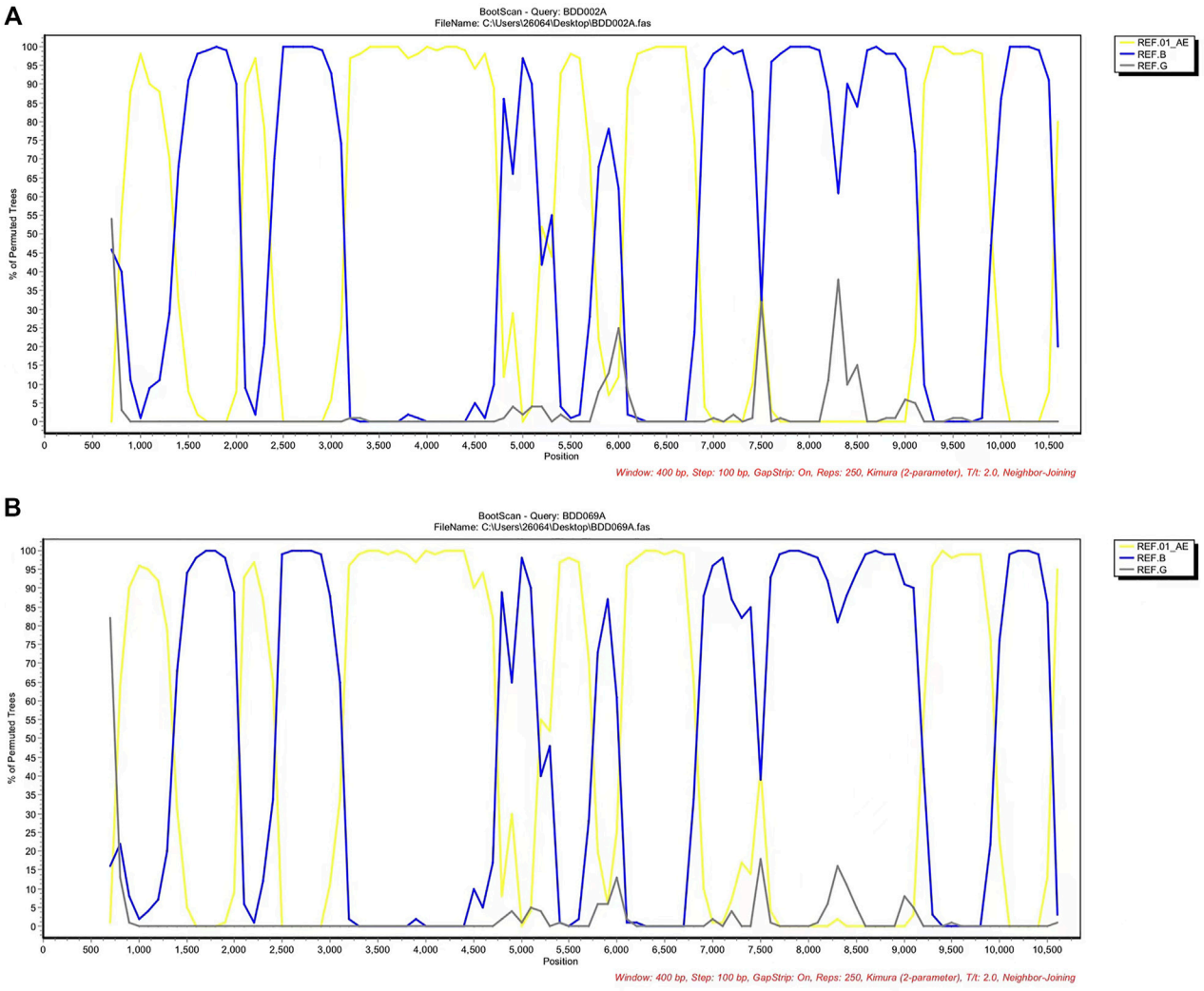


FIGURE 2
Bootscan results of the novel CRF01_AE/B identified. **(A)** Bootscan plots of BDD002A using CRF01_AE, subtype B and subtype G as references. The parameters of the bootscan analysis were a window size of 400 bp and a step size of 100 bp. **(B)** Bootscan plots of BDD069A using CRF01_AE and subtype B as putative parental reference sequences and subtype G as the outgroup. The parameters of the bootscan analysis were a window size of 400 bp and a step size of 100 bp. CRF, circulating recombinant form.

Case description

The two individuals, BDD002A and BDD069A, were a 35-year-old unmarried man and a 39-year-old married man, respectively. Moreover, their baseline $CD4^+$ T-cell counts were 242 cells/ul and 128 cells/ul, respectively while the HIV-1 viral load were 1,140,000 copies/ml and 703,000copies/ml, respectively. They were infected through homosexual transmission. The study was approved by the Medical Ethics Committee of the Baoding People's Hospital (protocol number: 2019-03). Written informed consent was obtained from the subjects prior to sample collection.

Diagnostic assessment

As described previously by us (Yang et al., 2022), HIV-1 RNA was extracted from 140 μL of the plasma samples of subjects BDD002A and BDD069A using a QIAamp Viral RNA Mini Accessory Set (QIAGEN, Hilden, Germany). PrimeScript IV 1st Strand cDNA Synthesis Mix (TaKaRa Biotechnology, Dalian, China) was used to reverse transcribe the RNA into 3′half-molecule cDNAs using the primers and reaction conditions listed in our previous report (Yang et al., 2022). Nested polymerase chain reactions (PCR) was performed using TaKaRa Premix Taq (TaKaRa Biotechnology, Dalian, China) to amplify 3′halfmolecule region of the NFLGs of BDD002A and BDD069A. HIV-1 near full-length *pol* and *gag* genes were amplified using Takara One-step RT-PCR Kit v2.0 (TaKaRa Biotechnology, Dalian, China). HIV-1 near full-length *pol* gene was amplified in two steps using the primers below:Pol-1e: TGGAAA TGTGGRAARGARGGAC (forward), Pol-x: CCTGTAATG CARMCCCCAATATG TT (reverse) and Pol-3: ACTGAGAGAC AG GCTAATTTTTTAGGGA (forward), Pol-4e:CTCCTAGTGG GGATRTGTACTTCTGARCTTA (reverse). HIV-1 near full-length *gag* gene was amplified in two steps using the primers below:gag-763:TGACTAGCGGAGGCTAGAAGG (forward), gag-5:TTCCYCC

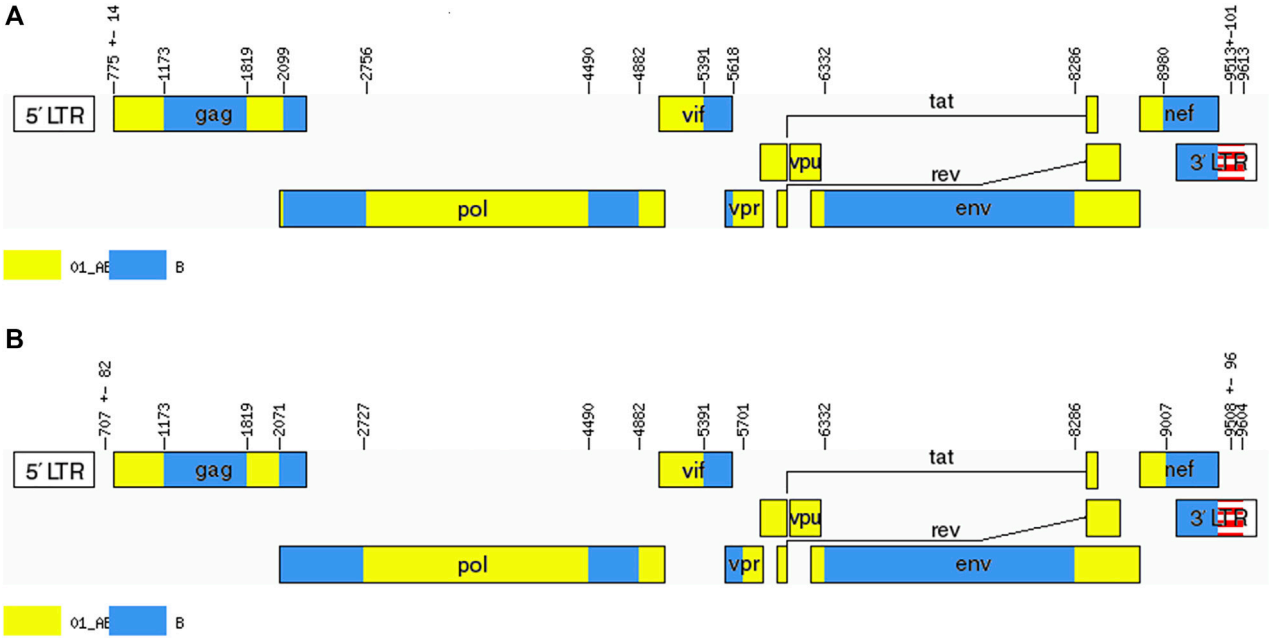


FIGURE 3
Recombinant maps of the novel CRF01_AE/B identified. The unique recombinant maps of **(A)** BDD002A and **(B)** BDD069A were drawn with the online Recombinant HIV-1 Drawing Tool (https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html). Six B segments were inserted into the CRF01_AE backbone, and 11 breakpoints divided the whole genome into 12 unique segments.

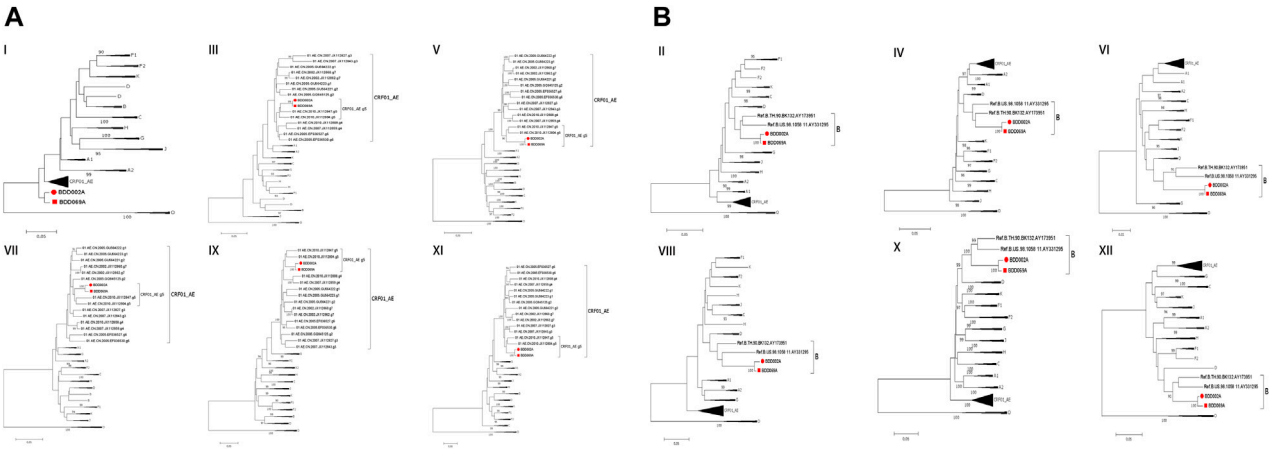


FIGURE 4
Subregional phylogenetic trees of the novel CRF01_AE/B identified. The trees were constructed by the neighbor-joining method with 1,000 bootstrap replications using Mega6.0. Bootstrap values ≥90% are shown at the corresponding nodes. The scale bars indicate a genetic distance of 5%. Each segment of BDD002A and BDD069A is marked by a red-filled circle and a red-filled square, respectively. **(A)** CRF01_AE regions for both URFs. **(B)** subtype B regions for both URFs.

TATCATT TTTGGTTTCC (reverse) and gag-617: TGTGGAAAAT CTCTAGCAGTGG (forward), gag-6:TAATGCTTTTATT TTYTCYT CTGTCAATGGC (reverse). The reaction conditions used for amplification have been reported previously (Lu et al., 2017; Fan et al., 2022a). The positive PCR products were purified using 1.0% agarose gel electrophoresis and sequenced using the Sanger sequencing technology by Tianyi Huiyuan Bioscience & Technology Inc. (Beijing, China). Two NFLGs were obtained by assembling them with near full-length *gag*, *pol* and 3′half-genome gene sequences using Sequencher 5.4.6 (Gene Codes Corp., Ann Arbor, MI, United States of America).

The NFLG sequences were then submitted to the online tool HIV BLAST (https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html) to determine whether the same recombinant sequences had been identified previously, but no sequences with high similarity (>95%) to BDD002A and BDD069A were found in the HIV database. In addition, phylogenetic tree and subregion phylogenetic trees were constructed using the neighbor-joining (N-J) method based on the

Kimura two-parameter model with 1,000 bootstrap replications by Mega6.0. The recombination pattern was determined by Recombination Identification Program (https://www.hiv.lanl.gov/content/sequence/RIP/RIP.html) and jpHMM. Recombination breakpoints were identified by SimPlot (v3.5.1.0) and Bootscan analysis.

We acquired two NFLG sequences with 8,810 bp (HXB2: 761–9,613) and 8,944 bp (HXB2: 625–9,604) from BDD002A and BDD069A, respectively. The constructed NFLG N-J tree showed that both BDD002A and BDD069A formed a monophyletic branch with a bootstrap value of 100%, indicating that BDD002A and BDD069A are two novel recombinant forms (Figure 1). The recombinant breakpoints analysis revealed that BDD002A and BDD069A were composed of 12 interlaced mosaic gene segments, including six CRF01_AE subregions (I, III, V, VII, IX and XI) and 6 B regions (II, IV, VI, VIII, X and XII), with 11 recombinant breakpoints relative to the HXB2 coordinate (Figure 2; Figure 3; Figure 4). Subregion analysis (Figure 3) confirmed that the gene mosaic structure of the two NFLGs are: ICRF01_AE (HXB2, 790–1,172 nt), IIB (HXB2, 1,173–1,818 nt), IIICRF01_AE (HXB2, 1,819–2,098 nt), IVB (HXB2, 2,099–2,755 nt), VCRF01_AE (HXB2, 2,756–4,489 nt), VIB (HXB2, 4,490–4,881 nt), VIICRF01_AE (HXB2, 4,882–5,390 nt), VIIIB (HXB2, 5,391–5,617 nt), IXCRF01_AE (HXB2, 5,618–6,331 nt), XB (HXB2, 6,332–8,285 nt), XICRF01_AE (HXB2, 8,286–8,979 nt) and XIIB (HXB2, 8,980–9,411 nt) for BDD002A (Figure 3A); and ICRF01_AE (HXB2, 790–1,172 nt), IIB (HXB2, 1,173–1,818 nt), IIICRF01_AE (HXB2, 1,819–2,070 nt), IVB (HXB2, 2,071–2,726 nt), VCRF01_AE (HXB2, 2,727–4,489 nt), VIB (HXB2, 4,490–4,881 nt), VIICRF01_AE (HXB2, 4,882–5,390 nt), VIIIB (HXB2, 5,391–5,700 nt), IXCRF01_AE (HXB2, 5,701–6,331 nt), XB (HXB2, 6,332–8,285 nt), XI CRF01_AE (HXB2, 8,286–9,006 nt) and XIIB (HXB2, 9,007–9,411 nt) for BDD069A (Figure 3B). The above data revealed that both NFLGs shared seven almost identical breakpoints except a minor difference within the *vif-vpr* gene region (Figure 3). The parental origin of all fragments of the two NFLGs were analyzed and the CRF01_AE regions for both URFs were from the CRF01_AE cluster five lineage, which is circulating primarily among MSM in major northern cities of China (Figure 4A) (Feng et al., 2013). The subtype B regions for both URFs were clustered within the northern China subtype B lineage, which also circulates primarily among MSM in northern China (Figure 4B) (Li et al., 2011).

Discussion

The diversity of HIV-1 is a significant challenge in preventing the global spread of HIV due to the different pathogenicity of subtypes. The co-circulation of multiple subtypes enables the occurrence of numerous URFs (Gao et al., 2021). In China, the prevalence of CRFs may be underestimated because of the high prevalence of URFs and the formation of many potential CRFs (Liu et al., 2019; Wang et al., 2021). URFs convert to CRFs if they spread widely and circulate in the population (Robertson et al., 2000; Hemelaar, 2012). In recent years, new CRFs such as CRF103_01 B and CRF112_01 B have been found because of the high occurrence of URFs among MSMs in Baoding (Zhou et al., 2020; Fan et al., 2022b; Wang et al., 2022). The continuous emergence of new recombinant forms has brought new barriers to the monitoring, treatment, vaccine development and prevention of HIV. In this study, we identified and characterized two novel recombinants derived from subtypes CRF01_AE and B, which were highly different from those reported previously in the cities of Langfang and Baoding and other Chinese provinces (Guo et al., 2014; Yan et al., 2015; Li et al., 2018; Huang et al., 2019; Li et al., 2019; Ou et al., 2019; Ge et al., 2020; Fan et al., 2022c). These two URFs were more complex than those reported previously. Although no new CRF has been formed, we infer that it may be a potential CRF. Currently, the sexual contact has been the main route of HIV-1 spread, it is key vital for us to carry out dynamic monitoring of new recombination forms in order to interdict HIV-1 spread.

In conclusion, we identified two almost identical HIV-1 URFs among MSMs in Baoding. The recombinant forms of CRF01_AE/B and CRF01_AE/CRF07_BC were found frequently in the MSM population of Baoding, indicating an increase in the genetic diversity of HIV-1 and the presence of many undiagnosed multiple infection cases in this region (Ou et al., 2019; Fan et al., 2022d; Yang et al., 2022), further bringing more barriers for HIV prevention and therapy. This study suggests that we should continuously monitor HIV-1 molecular epidemiology in order to provide effective suggestions to HIV-1 prevention and vaccine design.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, OP796643 https://www.ncbi.nlm.nih.gov/genbank/, OP796644.

Ethics statement

The studies involving human participants were reviewed and approved by The People's Hospital of Baoding, Baoding Ethics committee. The patients/participants provided their written informed consent to participate in this study.

Author contributions

BZ, XL and WA designed the study. JM, MS and WF acquired the sequences. BZ, SC, WA and XL analyzed and interpreted the data. BZ, SC, JM and MS wrote the manuscript. All authors read and agreed to the published version of the manuscript.

Acknowledgments

The authors thank all staff involved in this study. We thank Liwen Bianji (Edanz) (www.liwenbianji.cn) for editing the English text of a draft of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Fan, W., Liu, Y., Li, Y., Su, M., Meng, J., Lu, X., et al. (2022). Identification of three novel HIV-1 second-generation recombinant forms (CRF01_AE/CRF07_BC) among men who have sex with men in baoding, Hebei, China. *AIDS Res. Hum. Retroviruses* 38 (10), 812–816. doi:10.1089/AID.2022.0037

Fan, W., Su, M., Meng, J., Yang, X., Liu, Z., Wang, H., et al. (2022). Characterization of a new HIV-1 CRF01_AE/B recombinant virus form among men who have sex with men in baoding, Hebei, China. *AIDS Res. Hum. Retroviruses* 38 (3), 237–241. doi:10.1089/AID.2021.0209

Fan, W., Xing, Y., Han, L., Su, M., Meng, J., Dai, E., et al. (2022). HIV-1 genetic characteristics and pre-treatment drug resistance among newly diagnosed population in Baoding, Hebei Province [J]. *Chin. J. Microbiol. Immunol.* 42 (02), 88–93. doi:10.3760/cma.j.cn112309-20210814-00272

Fan, W., Yang, X., Liu, Z., Su, M., Meng, J., Lu, X., et al. (2022). Subtypes and transmission network analysis of HIV-1 among men who have sex with men in Baoding [J]. *Int. J. Virology* 29 (01), 30–34. doi:10.3760/cma.j.issn.1673-4092.2022.01.006

Feng, Y., He, X., His, J. H., Li, F., Li, X., Wang, Q., et al. (2013). The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. *AIDS* 27 (11), 1793–1802. doi:10.1097/QAD.0b013e328360db2d

Gao, Y., He, S., Tian, W., Li, D., An, M., Zhao, B., et al. (2021). First complete-genome documentation of HIV-1 intersubtype superinfection with transmissions of diverse recombinants over time to five recipients. *PLoS Pathog.* 17 (2), e1009258. doi:10.1371/journal.ppat.1009258

Ge, Z., Getaneh, Y., Liang, Y., Lv, B., Liu, Z., Li, K., et al. (2020). Identification of a novel HIV-1 second-generation (CRF01_AE/B) among men who have sex with men in Tianjin, China. *AIDS Res. Hum. Retroviruses* 36 (2), 138–142. doi:10.1089/AID.2019.0187

Guo, H., Hu, H., Zhou, Y., Yang, H., Huan, X., Qiu, T., et al. (2014). A Novel HIV-1 CRF01_AE/B recombinant among men who have sex with men in Jiangsu Province, China. *AIDS Res. Hum. Retroviruses* 30 (7), 706–710. doi:10.1089/AID.2014.0012

He, X., Xing, H., Ruan, Y., Hong, K., Cheng, C., Hu, Y., et al. (2012). A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. *PLoS One* 7 (10), e47289. doi:10.1371/journal.pone.0047289

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18 (3), 182–192. doi:10.1016/j.molmed.2011.12.001

Huang, Q., Ou, W., Feng, Y., Li, F., Li, K., Sun, J., et al. (2019). Near full-length genomic characterization of HIV-1 CRF01_AE/B recombinant strains identified in Hebei, China. *AIDS Res. Hum. Retroviruses* 35 (2), 196–204. doi:10.1089/AID.2018.0130

Li, K., Liu, M., Li, J., Dong, A., Zhou, Y., Ding, Y., et al. (2019). Genomic characterization of a novel HIV-1 second-generation recombinant form (CRF01_AE/B) from men who have sex with men in guangxi Zhuang autonomous region, China. *AIDS Res. Hum. Retroviruses* 35 (10), 972–977. doi:10.1089/AID.2019.0149

Li, K., Ou, W., Feng, Y., Sun, J., Ge, Z., Xing, H., et al. (2018). Near full-length genomic characterization of a novel HIV type 1 recombinant form (CRF01_AE/B) identified from anhui, China. *AIDS Res. Hum. Retroviruses* 34 (12), 1100–1105. doi:10.1089/AID.2018.0144

Li, L., Lu, X., Li, H., Chen, L., Wang, Z., Liu, Y., et al. (2011). High genetic diversity of HIV-1 was found in men who have sex with men in Shijiazhuang, China. *Infect. Genet. Evol.* 11 (6), 1487–1492. doi:10.1016/j.meegid.2011.05.017

Liu, Y., Su, B., Zhang, Y., Jia, L., Li, H., Li, Z., et al. (2019). Brief report: Onward transmission of multiple HIV-1 unique recombinant forms among men who have sex with men in beijing, China. *J. Acquir Immune Defic. Syndr.* 81 (1), 1–4. doi:10.1097/QAI.0000000000001983

Lu, X., Kang, X., Liu, Y., Cui, Z., Guo, W., Zhao, C., et al. (2017). HIV-1 molecular epidemiology among newly diagnosed HIV-1 individuals in Hebei, a low HIV prevalence province in China. *PLoS One* 12 (2), e0171481. doi:10.1371/journal.pone.0171481

Lu, X., Zhang, J., Wang, Y., Liu, M., Li, Y., An, N., et al. (2020). Large transmission clusters of HIV-1 main genotypes among HIV-1 individuals before antiretroviral therapy in the Hebei province, China. *AIDS Res. Hum. Retroviruses* 36 (5), 427–433. doi:10.1089/AID.2019.0199

Moore, M. D., and Hu, W. S. (2009). HIV-1 RNA dimerization: It takes two to tango. *AIDS Rev.* 11 (2), 91–102.

Ou, W., Li, K., Feng, Y., Huang, Q., Ge, Z., Sun, J., et al. (2019). Characterization of a new HIV-1 CRF01_AE/B recombinant virus form among men who have sex with men in shanghai, China. *AIDS Res. Hum. Retroviruses* 35 (4), 414–418. doi:10.1089/AID.2018.0197

Robertson, D. L., Anderson, J. P., Bradac, J. A., Carr, J. K., Foley, B., Funkhouser, R. K., et al. (2000). HIV-1 nomenclature proposal. *Science* 288 (5463), 55–56. doi:10.1126/science.288.5463.55d

Shi, P., Chen, Z., Meng, J., Su, M., Yang, X., Fan, W., et al. (2021). Molecular transmission networks and pre-treatment drug resistance among individuals with acute HIV-1 infection in Baoding, China. *PLoS One* 16 (12), e0260670. doi:10.1371/journal.pone.0260670

Shi, P., Wang, X., Fan, W., Zhang, Z., Su, M., Meng, J., et al. (2021). Pre-treatment drug resistance analysis of HIV-1 infected patients in Baoding City, 2019-2020. *Chin. J. Dermatovenereology* 35 (09), 1012–1016+ 1020. doi:10.13735/j.cjdv.1001-7089.202103052

Wang, H., Zhao, X., Su, M., Meng, J., Fan, W., and Shi, P. (2022). Identification of a new HIV-1 circulating recombinant form CRF112_01B strain in baoding city, Hebei province, China. *Curr. HIV Res.* 20, 485–491. doi:10.2174/1570162X21666221027122528

Wang, T. X. (2020). A summary of HIV/AIDS epidemic situation across the country in 2020. Available at https://xw.qq.com/amphtml/20201207A0B1EE00 (accessed December 7, 2020).

Wang, X., Zhang, Y., Liu, Y., Li, H., Jia, L., Han, J., et al. (2021). Phylogenetic analysis of sequences in the HIV Database revealed multiple potential circulating recombinant forms in China. *AIDS Res. Hum. Retroviruses* 37 (9), 694–705. doi:10.1089/AID.2020.0190

Yan, J., Xin, R., Li, Z., Feng, Y., Lu, H., Liao, L., et al. (2015). CRF01_AE/B/C, a novel drug-resistant HIV-1 recombinant in men who have sex with men in beijing, China. *AIDS Res. Hum. Retroviruses* 31 (7), 745–748. doi:10.1089/AID.2015.0064

Yang, X., Zhu, H., An, W., Zhao, J., Lu, X., Sun, W., et al. (2022). Genetic characterizationof a novel HIV-1 CRF01_AE/CRF07_BC recombinant form found among men who have sex with men in Baoding City, Hebei Province, China. *Arch. Virol.* 167 (11), 2395–2402. doi:10.1007/s00705-022-05563-y

Yebra, G., Frampton, D., Gallo, C. T., Raffle, J., Hubb, J., Ferns, R. B., et al. (2018). A high HIV-1 strain variability in London, UK, revealed by full-genome analysis: Results from the ICONIC project. *PLoS One* 13 (2), e0192081. doi:10.1371/journal.pone.0192081

Zhou, J., Lu, X., Feng, Y., Li, M., Zhu, Y., Kang, R., et al. (2020). Genome sequence of a novel HIV-1 circulating recombinant form (CRF103_01B) identified from Hebei province, China. *AIDS Res. Hum. Retroviruses* 36 (3), 234–241. doi:10.1089/AID.2019.0196

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 03 May 2023
DOI 10.3389/fmicb.2023.1159928

OPEN ACCESS

EDITED BY
Binhua Liang,
Public Health Agency of Canada (PHAC), Canada

REVIEWED BY
Teiichiro Shiino,
National Center For Global Health and Medicine, Japan
Marcel Tongo,
Center for Research on Emerging and Re-Emerging Diseases (CREMER), Cameroon

*CORRESPONDENCE
Weiguang Fan
✉ fwggwf@outlook.com

†These authors have contributed equally to this work

RECEIVED 08 February 2023
ACCEPTED 10 April 2023
PUBLISHED 03 May 2023

CITATION
Yang X, Zhao N, Su M, Meng J, Du J, An W, Shi H and Fan W (2023) Characterization of two novel HIV-1 second-generation recombinants (CRF01_AE/CRF07_BC) identified in Hebei Province, China.
Front. Microbiol. 14:1159928.
doi: 10.3389/fmicb.2023.1159928

Characterization of two novel HIV-1 second-generation recombinants (CRF01_AE/CRF07_BC) identified in Hebei Province, China

Xuegang Yang[1†], Na Zhao[1†], Miaomiao Su[1†], Juan Meng[1], Jian Du[1], Weina An[2], Haoxi Shi[2] and Weiguang Fan[2*]

[1]Infection Division, The People's Hospital of Baoding, Baoding, Hebei, China, [2]Clinical Laboratory, The People's Hospital of Baoding, Baoding, Hebei, China

Introduction: The unique recombinant forms (URFs) of HIV-1 consist of a mixture of subtypes, and each URF has a unique breakpoint. In this study, we identified the near fulllength genome (NFLG) sequences of two novel HIV-1 URFs (Sample ID: BDD034A and BDL060) isolated during HIV-1 molecular surveillance in 2022 in Baoding city, Hebei Province, China.

Methods: The two sequences were aligned with subtype reference sequences and CRFs from China using MAFFT v7.0, and the alignments were adjusted manually using BioEdit (v7.2.5.0). Phylogenetic and subregion trees were constructed using MEGA11 with the neighbor-joining (N-J) method. Recombination breakpoints were identified by SimPlot (v3.5.1) based on Bootscan analyses.

Results: Recombinant breakpoint analysis revealed that the NFLGs of BDD034A and BDL060 were composed of CRF01_AE and CRF07_BC, containing seven segments, respectively. For BDD034A, three CRF01_AE fragments were inserted into the CRF07_BC main framework, whereas for BDL060, three CRF07_BC fragments were inserted into the CRF01_AE main framework.

Discussion: The emergence of the CRF01_AE/CRF07_BC recombinant strains indicates that HIV-1 co-infection is common. The increasing genetic complexity of the HIV-1 epidemic in China warrants continued investigation.

KEYWORDS

HIV, circulating recombinant forms, near full-length genome, unique recombination forms, MSM

Introduction

The high rates of mutation, recombination and replication that are characteristic of HIV-1 mean that new circulating recombinant forms (CRFs) and unique recombinant forms (URFs) are constantly emerging (Hemelaar, 2012; Bbosa et al., 2019). To date, 132 CRFs, as well as numerous URFs, have been registered in the Los Alamos National Laboratory HIV database.[1] In China, CRF01_AE and CRF07_BC are the predominant intersubtype

1 https://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html

TABLE 1 Epidemiological information about the two participants.

Strain name	Sex	Age (years)	Marital status	Transmission route	CD4 T-cell count (cells/μL)	HIV-1 viral load (copies/mL)	Accession number
BDD034A	Male	44	Married	MSM	571	35,200	OP745422
BDL060	Male	45	Married	MSM	379	43,000	OQ207706

recombinant forms of HIV-1, with recombinant forms between these two subtypes emerging, most prevalently among men who have sex with men (MSM), in recent years (He et al., 2012; Li et al., 2016; Yin et al., 2019). The continuous emergence of new recombinant forms has brought new challenges to the monitoring, treatment and prevention of HIV infection. Hebei is a northern province of China with a low HIV prevalence (Lu et al., 2017); however, over the last 10 years, the number of individuals infected with HIV-1 through sexual contact has reached 98.9%, among which 77.5% of cases are in MSM, with CRF01_AE (49.6%), CRF07_BC (29.7%) and B subtype (13.0%) being the three main genotypes (Lu et al., 2020). The co-circulation of CRF01_AE and CRF07_BC strains and the dual infection of these strains in the sexually-active population has increased the generation of inter-subtype recombinant forms (Luan et al., 2017). In this study, we detected and characterized 2 second generation HIV-1 recombinant strains (BDD034A and BDL060) derived from the CRF01_AE and CRF07_BC subtypes, isolated from MSM infected with HIV-1.

Materials and methods

In this study, we identified the near full-length genome (NFLG) sequences of two novel HIV-1 URF strains (Sample ID: BDD034A and BDL060) isolated during HIV-1 molecular surveillance in 2022 in Baoding city, Hebei Province, China. The two individuals from which these strains were isolated, BDD034A and BDL060, were a 44-year-old unmarried man and a 45-year-old married man, respectively (see Table 1 for further details). This study was approved by the Medical Ethics Committee of Baoding People's Hospital (protocol number: 2019–03). Written informed consent was obtained from the subjects prior to sample collection.

RNA was extracted from 140 μL of each subject's plasma sample using a QIAamp Viral RNA Mini Kit (Qiagen, Duesseldorf, Germany) in accordance with the manufacturer's instructions. PrimeScript IV 1st Strand cDNA Synthesis Mix (TaKaRa Biotechnology, Dalian, China) was used to reverse transcribe the RNA into 3′ and 5′ half-molecule cDNAs using the primers 1.R3. B3R: 5′-ACTACTTGAAG CACTCAAGGCAAGC TTTATTG-3′ and 07Rev8: 5′-CCTART GGGATGTGTACTT CTG AACTT-3′. A nested polymerase chain reaction (PCR) was performed using TaKaRa Premix Taq (TaKaRa Biotechnology) to amplify the 3′ and 5′ half-molecule regions of the NFLG sequences of BDD034A and BDL060. The reaction conditions and primer sequences used for amplification have been reported previously (Yang et al., 2022). PCR products were detected using 1.0% agarose gel electrophoresis, and amplified products of the expected size were purified from the corresponding electrophoretic bands and sequenced using Sanger sequencing technology by Tianyi Huiyuan Bioscience & Technology Inc. (Beijing, China).

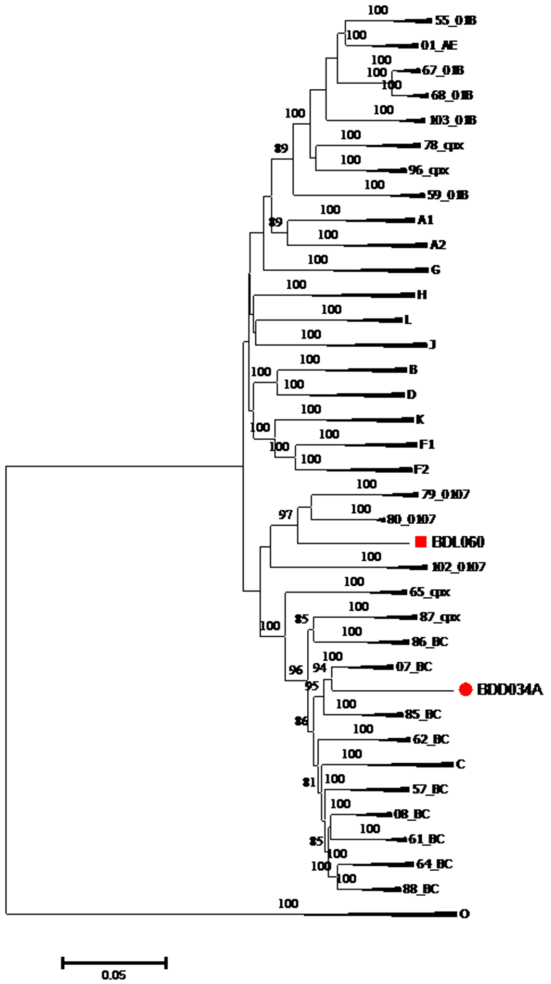


FIGURE 1
Phylogenetic tree based on the NFLG sequences. The neighbor-joining tree of BDD034A and BDL060 was constructed using Mega 11 with a bootstrap value of 1,000 replicates. Only bootstrap values ≥75% are shown in the corresponding nodes. The scale bar represents a 5% genetic distance.

The two NFLG sequences were submitted to the online tool HIV BLAST[2] to search for similar sequences. Then the two sequences were aligned with subtype reference sequences and CRFs from China[3] using MAFFT v7.0 (Katoh and Standley, 2013), and the alignments were adjusted manually using BioEdit (v7.2.5.0). Phylogenetic and subregion trees were constructed

2 https://www.hiv.lanl.gov/content/sequence/BASIC_BLAST/basic_blast.html
3 https://hiv.lanl.gov/components/sequence/HIV/search/ search.html

using MEGA11 (Tamura et al., 2021) with the neighbor-joining (N-J) method (Saitou and Nei, 1987). Recombination breakpoints were identified by SimPlot (v3.5.1) based on Bootscan analyses.

Results

We acquired two NFLG sequences of 8,890 bp (HXB2: 723–9,613) and 8,847 bp (HXB2: 757–9,604) from BDD034A and BDL060,

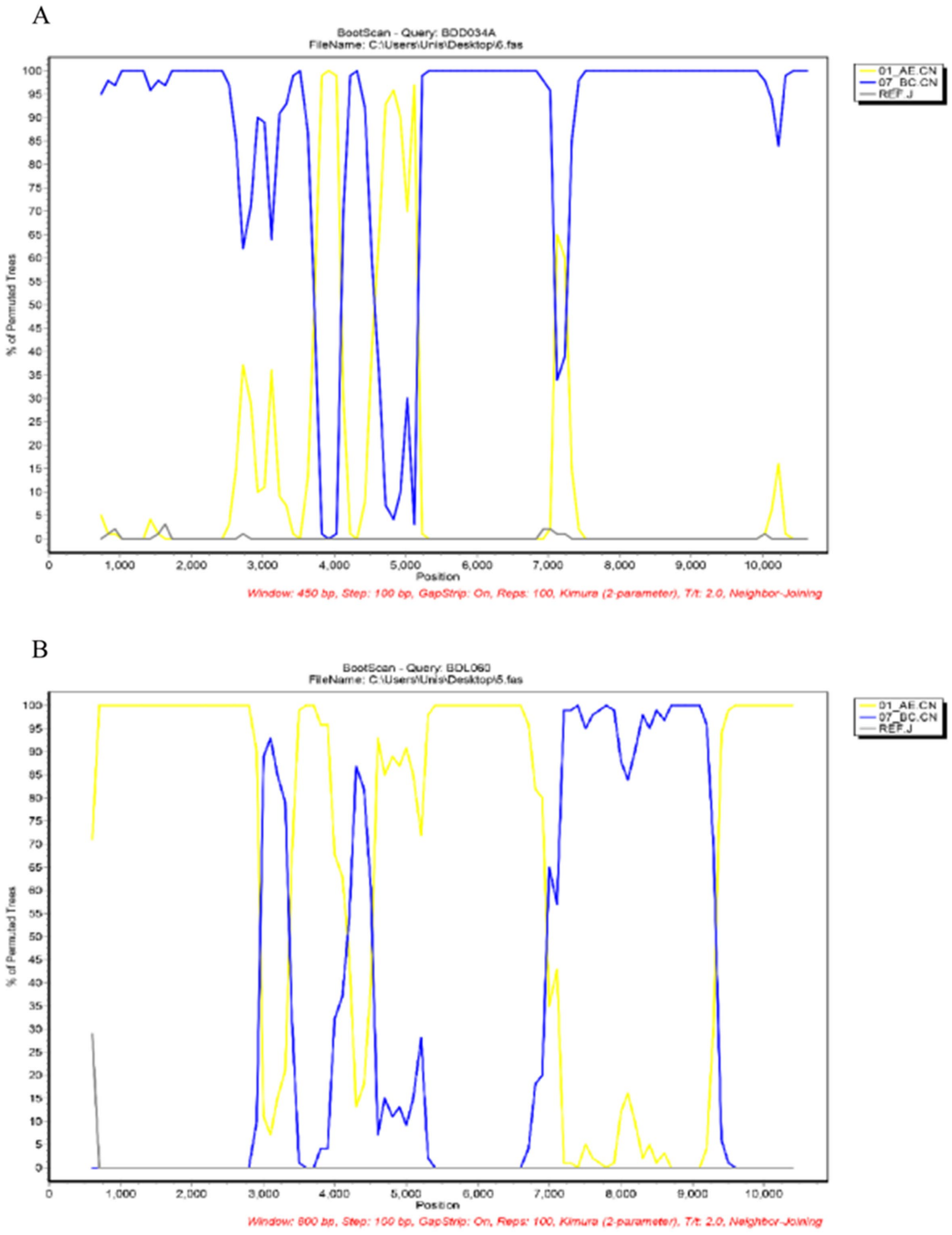


FIGURE 2
Bootscan analysis. **(A)** Bootscan plot of BDD034A using CRF01_AE (accession numbers JX112801 and JX112859), CRF07_BC (accession numbers HQ215552 and AF286226) and J (accession numbers EF614151 and AF082394) as reference sequences. The parameters were set to a window size of 800 and a step size of 100. **(B)** Bootscan plot of BDL060. The reference sequences were the same as those for BDD034A above. The parameters were set to a window size of 450 and a step size of 100.

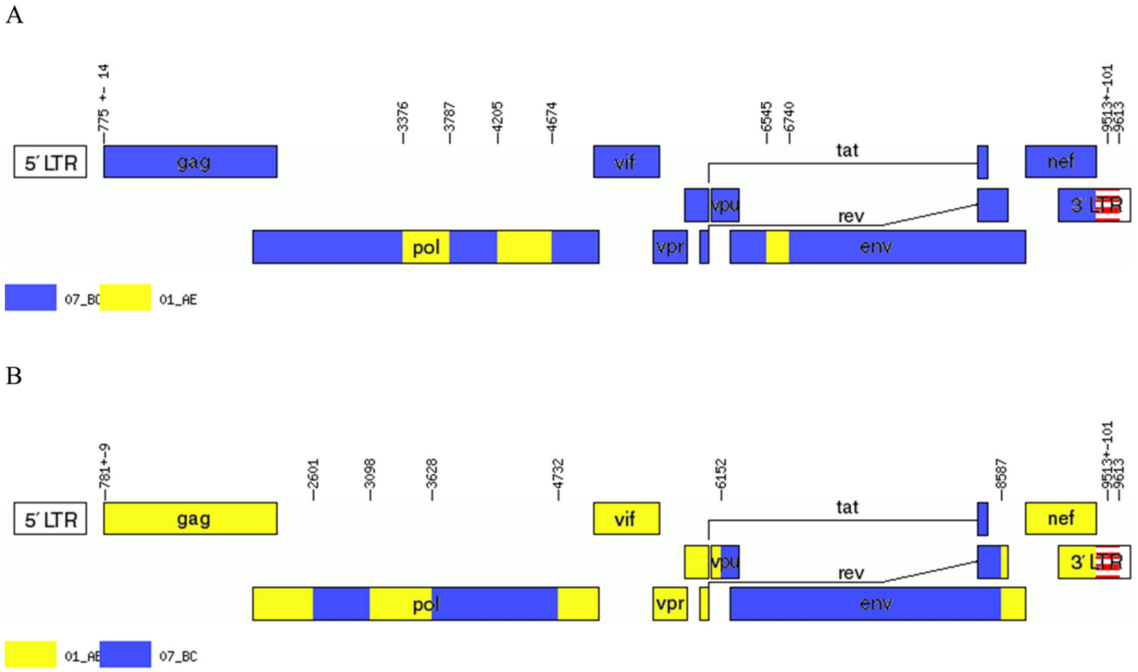


FIGURE 3
Genetic map of BDD034A **(A)** and BDL060 **(B)**. The Recombinant HIV-1 Drawing Tool was used, which is available at the HIV database: https://www.hiv.lanl.gov/content/sequence/DRAW_CRF/recom_mapper.html.

respectively. The results of HIV BLAST suggested that there are no ≥95% similar sequences found in the HIV database. The constructed NFLG N-J tree showed that both BDD034A and BDL060 formed separate monophyletic branches, indicating that BDD034A and BDL060 are two different novel recombinant forms (Figure 1). The recombinant breakpoint analysis revealed that BDD034A and BDL060 were composed of seven interleaved mosaic gene fragments (Figures 2, 3). However, the difference between these sequences was that BDD034A is a combination of three CRF01_AE fragments inserted into the CRF07_BC main framework, whereas BDL060 is a combination of three CRF07_BC fragments inserted into the CRF01_AE main framework (Figure 3A and Figure 3B). The mosaic recombinant structure of the BDD034A sequence can be described as follows: ICRF07_BC (HXB2, 790–3,375 nt); IICRF01_AE (HXB2, 3,376–3,786 nt); IIICRF07_BC (HXB2, 3,787–4,204 nt); IVCRF01_AE (HXB2, 4,205–4,673 nt); VCRF07_BC (HXB2, 4,674–6,544 nt); VICRF01_AE (HXB2, 6,545–6,739 nt); and VIICRF07_BC (HXB2, 6,740–9,411 nt). Subregion phylogenetic analysis was used to confirm the genetic origin of each segment. Segments I, III, V and VII were clustered with CRF07_BC, which is prevalent among MSM in northern China (Figure 4A). Segment II was clustered with CRF01_AE cluster 4 and segment VI, segment IV was clustered with CRF01_AE cluster 5 (Figure 4A). The mosaic recombinant structure of the BDL060 sequence can be described as follows: ICRF01_AE (HXB2, 790–2,600 nt); IICRF07_BC (HXB2, 2,601–3,097 nt); IIICRF01_AE (HXB3, 098–3,627 nt); IVCRF07_BC (HXB3, 628–4,731 nt); VCRF01_AE (HXB2, 4,732–6,151 nt); VICRFF07_BC (HXB2, 6,152–8,587 nt); and VIICRF01_AE (HXB2, 8,588–9,411 nt). Subregion phylogenetic analyses indicated that four CRF01_AE segments mainly originated from CRF01_AE cluster 4 strains among the MSM population in China (Figure 4B).

Discussion

CRF01_AE is the most important HIV-1 strain in China owing to its significant contribution to the HIV-1 epidemic, and it has been found to possess at least seven gene clusters. CRF01_AE cluster 1 was found primarily among heterosexuals and intravenous drug users (IDUs) in the southern provinces of China. CRF01_AE clusters 1, 2 and 3 were prevalent among heterosexuals and IDUs in southern and southwestern provinces of China. CRF01_AE clusters 4 and 5 were mainly distributed among the MSM population in northern China, including Beijing and Tianjin. Clusters 6 and 7 were only detected among heterosexuals in two southeast and southwest provinces (Feng et al., 2013; Li et al., 2017). In this study, phylogenetic analysis of BDD034A showed that segments II, IV and VI clustered within clusters 4 and 5 of CRF01_AE, respectively. Because BDD034A showed intra-subtype recombination, it does not rule out the possibility of dual infection. CRF07_BC is another dominant HIV-1 strain in China that was originally isolated in 1993 in an IDU in Yunnan Province (Meng et al., 2012), and then spread along drug trafficking routes to Sichuan, Guangxi and Xinjiang Provinces (Tee et al., 2008) in southwestern and northern China. At present, CRF01_AE and CRF07_BC are the predominant intersubtype recombinants among sexually-active populations, especially among MSM in China (Li et al., 2016). High geographical mobility, condom-free sex and multiple sexual partners are contributing factors that lead to an increase in CRFs of HIV-1. The extremely high-risk behaviors and cocirculation of multiple subtypes make MSM significantly more vulnerable to dual infection (Liu et al., 2019). According to our previous report

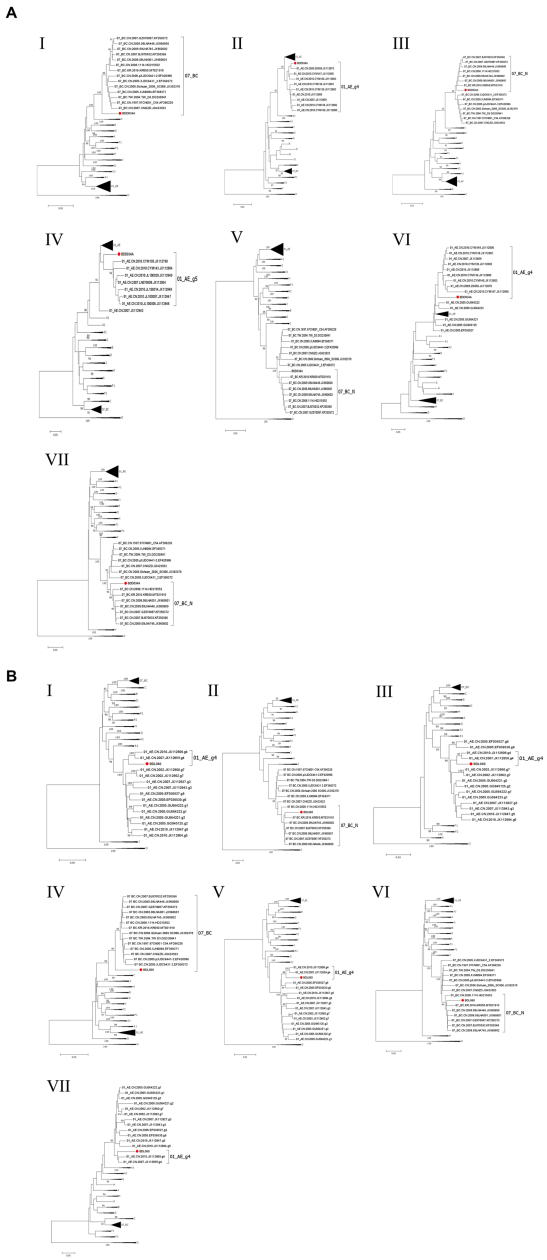


FIGURE 4
The subregion trees of BDD034A **(A)** and BDL060 **(B)**. The subregion phylogenetic trees were constructed using Mega 11 software and the neighbor-joining method with 1,000 bootstrap replications. Bootstrap values ≥90% are shown at the corresponding nodes. The scale bars represent a 5% genetic distance. Segments of BDD034A and BDL060 are indicated by a solid red circle and a solid red square, respectively.

(Shi et al., 2021), the educational level of MSM in Baoding City is low, and they have limited knowledge of HIV prevention. Both of BDD034A and BDL060 have primary education, multiple sexual partners and unprotected anal intercourse. Future research should focus on MSM with low educational level, learn more about HIV prevention, strengthen self-protection, and reduce the risk of HIV transmission. Hebei Province is located in northern China, surrounding Beijing and Tianjin. The convenient transport network in this region enables geographical mobility, and creates opportunities for dual or multiple infections within the MSM population. This has led to the emergence and prevalence of new recombinant strains in MSM in recent years, such as CRF103_01B, CRF123_0107, CRF01_AE/B and CRF01_AE/CRF07_BC (Huang et al., 2019; Zhou et al., 2020; Han et al., 2021; Xing et al., 2021; Fan et al., 2022a,b; Xing et al., 2022). The emergence of new recombinant forms have increased the diversity of HIV-1 isolates prevalent in Hebei province, indicating the further molecular monitoring of HIV-1 diversity is vital in the region.

Conclusion

In conclusion, we identified two novel recombinant forms of HIV-1 isolated from MSM, which have no similar breakpoints from the CRFs and URFs reported previously. The BLAST search results indicated there are no ≥95% similar sequences in the database with the BDD034A and BDL060 sequences. The emergence of CRF01_AE and CRF07_BC recombinant forms might suggest high genetic variation among HIV-1 in Hebei, warning us to continuously supervise HIV-1 molecular epidemiologic dynamics and gather enough information for vaccine design and to provide effective suggestions for accurate control.

Data availability statement

The nucleotide sequences of BDD034A and BDL060 have been deposited in the NCBI GenBank database under accession numbers OP745422 and OQ207706, respectively.

Ethics statement

The studies involving human participants were reviewed and approved by the Medical Ethics Committee at the People's Hospital of Baoding (2019–03). Written informed consent to participate in this study was provided by the patient/participant or the legal guardian/next of kin of the patient/participant.

Author contributions

XY, NZ, MS, and WF designed the study. HS, JM, and WA acquired the sequences. JM, JD, MS, and WF analyzed and interpreted the data. XY, NZ, and MS wrote the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors thank all staff involved in this study. We also thank Liwen Bianji (Edanz) (www.liwenbianji.cn) for editing the English text of a draft of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1159928/full#supplementary-material

References

Bbosa, N., Kaleebu, P., and Ssemwanga, D. (2019). HIV subtype diversity worldwide. *Curr. Opin. HIV AIDS* 14, 153–160. doi: 10.1097/COH.0000000000000534

Fan, W., Liu, Y., Li, Y., Su, M., Meng, J., Lu, X., et al. (2022a). Identification of three novel HIV-1 second-generation recombinant forms (CRF01_AE/CRF07_BC) among men who have sex with men in Baoding, Hebei, China. *AIDS Res. Hum. Retrovir.* 38, 812–816. doi: 10.1089/AID.2022.0037

Fan, W., Su, M., Meng, J., Yang, X., Liu, Z., Wang, H., et al. (2022b). Characterization of a new HIV-1 CRF01_AE/B recombinant virus form among men who have sex with men in Baoding, Hebei, China. *AIDS Res. Hum. Retrovir.* 38, 237–241. doi: 10.1089/AID.2021.0209

Feng, Y., He, X., Hsi, J. H., Li, F., Li, X., Wang, Q., et al. (2013). The rapidly expanding CRF01_AE epidemic in China is driven by multiple lineages of HIV-1 viruses introduced in the 1990s. *AIDS* 27, 1793–1802. doi: 10.1097/QAD.0b013e328360db2d

Han, L., Li, H., Wang, L., Jia, L., Han, J., Li, T., et al. (2021). Near full-length genomic characterization of two novel HIV-1 unique recombinant forms (CRF01_AE/CRF07_BC) among men who have sex with men in Shijiazhuang City, Hebei Province, China. *AIDS Res. Hum. Retrovir.* 37, 978–984. doi: 10.1089/AID.2021.0100

He, X., Xing, H., Ruan, Y., Hong, K., Cheng, C., Hu, Y., et al. (2012). A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. *PLoS One* 7:e47289. doi: 10.1371/journal.pone.0047289

Hemelaar, J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182–192. doi: 10.1016/j.molmed.2011.12.001

Huang, Q., Ou, W., Feng, Y., Li, F., Li, K., Sun, J., et al. (2019). Near full-length genomic characterization of HIV-1 CRF01_AE/B recombinant strains identified in Hebei, China. *AIDS Res. Hum. Retrovir.* 35, 196–204. doi: 10.1089/AID.2018.0130

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Li, X., Li, W., Zhong, P., Fang, K., Zhu, K., Musa, T. H., et al. (2016). Nationwide trends in molecular epidemiology of HIV1 in China. *AIDS Res. Hum. Retrovir.* 32, 851–859. doi: 10.1089/AID.2016.0029

Li, X., Liu, H., Liu, L., Feng, Y., Kalish, M. L., Ho, S. Y. W., et al. (2017). Tracing the epidemic history of HIV-1 CRF01_AE clusters using near-complete genome sequences. *Sci. Rep.* 7:4024. doi: 10.1038/s41598-017-03820-8

Liu, Y., Su, B., Zhang, Y., Jia, L., Li, H., Li, Z., et al. (2019). Brief report: onward transmission of multiple HIV-1 unique recombinant forms among men who have sex with men in Beijing. *China J. Acquir. Immune Defic. Syndr.* 81, 81, 1–4. doi: 10.1097/QAI.0000000000001983

Lu, X., Kang, X., Liu, Y., Cui, Z., Guo, W., Zhao, C., et al. (2017). HIV-1 molecular epidemiology among newly diagnosed HIV-1 individuals in Hebei, a low HIV prevalence province in China. *PLoS One* 12:e0171481. doi: 10.1371/journal.pone.0171481

Lu, X., Zhang, J., Wang, Y., Liu, M., Li, Y., An, N., et al. (2020). Large transmission clusters of HIV-1 Main genotypes among HIV-1 individuals before antiretroviral therapy in the Hebei Province. *China AIDS Res. Hum. Retroviruses* 36, 427–433. doi: 10.1089/AID.2019.0199

Luan, H., Han, X., Yu, X., An, M., Zhang, H., Zhao, B., et al. (2017). Dual infection contributes to rapid disease progression in men who have sex with men in China. *J. Acquir. Immune Defic. Syndr.* 75, 480–487. doi: 10.1097/QAI.0000000000001420

Meng, Z., Xin, R., Zhong, P., Zhang, C., Abubakar, Y. F., Li, J., et al. (2012). A new migration map of HIV-1 CRF07_BC in China: analysis of sequences from 12 provinces over a decade. *PLoS One* 7:e52373. doi: 10.1371/journal.pone.0052373

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Shi, P., Wang, X., Fan, W., et al. (2021). Pre-treatment drug resistance analysis of HIV-1 infected patients in Baoding City, 2019– 2020. *Chinese J Dermatovenereol.* 35, 1012–1016+1020. doi: 10.13735/j.cjdv.1001-7089.202103052

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tee, K. K., Pybus, O. G., Li, X. J., Han, X., Shang, H., Kamarulzaman, A., et al. (2008). Temporal and spatial dynamics of human immunodeficiency virus type 1 circulating recombinant forms 08_BC and 07_BC in Asia. *J. Virol.* 82, 9206–9215. doi: 10.1128/JVI.00399-08

Xing, Y., Guo, Y., Wang, L., Li, H., Han, J., Wang, X., et al. (2021). Identification of two novel HIV-1 second-generation recombinant forms (CRF01_AE/CRF07_BC) in Hebei, China. *AIDS Res. Hum. Retrovir.* 37, 967–972. doi: 10.1089/AID.2021.0057

Xing, Y., Wang, L., Li, Y., Wang, Y., Han, L., Huang, G., et al. (2022). Identification of a new HIV-1 intersubtype circulating recombinant form (CRF123_0107) in Hebei province, China. *J. Infect.* 84, e36–e39. doi: 10.1016/j.jinf.2022.01.017

Yang, X., Zhu, H., An, W., Zhao, J., Lu, X., Sun, W., et al. (2022). Genetic characterization of a novel HIV-1 CRF01_AE/ CRF07_BC recombinant form found among men who have sex with men in Baoding City, Hebei Province. *China Arch Virol.* 167, 2395–2402. doi: 10.1007/s00705-022-05563-y

Yin, Y., Liu, Y., Zhu, J., Hong, X., Yuan, R., Fu, G., et al. (2019). The prevalence, temporal trends, and geographical distribution of HIV-1 subtypes among men who have sex with men in China: a systematic review and meta-analysis. *Epidemiol. Infect.* 147:e83. doi: 10.1017/S0950268818003400

Zhou, J., Lu, X., Feng, Y., Li, M., Zhu, Y., Kang, R., et al. (2020). Genome sequence of a novel HIV-1 circulating recombinant form (CRF103_01B) identified from Hebei Province, China. *AIDS Res. Hum. Retrovir.* 36, 234–241. doi: 10.1089/AID.2019.0196

frontiers | Frontiers in Microbiology

TYPE Review
PUBLISHED 25 July 2023
DOI 10.3389/fmicb.2023.1228128

OPEN ACCESS

EDITED BY
Hezhao Ji,
Public Health Agency of Canada, Canada

REVIEWED BY
Jasdeep Singh,
University of Denver, United States
Xiang Gao,
Loyola University Chicago, United States

*CORRESPONDENCE
Guangwen Cao
✉ gcao@smmu.edu.cn

RECEIVED 24 May 2023
ACCEPTED 03 July 2023
PUBLISHED 25 July 2023

CITATION
Fang L, Xu J, Zhao Y, Fan J, Shen J, Liu W and Cao G (2023) The effects of amino acid substitution of spike protein and genomic recombination on the evolution of SARS-CoV-2.
Front. Microbiol. 14:1228128.
doi: 10.3389/fmicb.2023.1228128

The effects of amino acid substitution of spike protein and genomic recombination on the evolution of SARS-CoV-2

Letian Fang[1,2,3], Jie Xu[4], Yue Zhao[1,2,3], Junyan Fan[1,2,3], Jiaying Shen[5], Wenbin Liu[1,2,3] and Guangwen Cao[1,2,3]*

[1]Key Laboratory of Biological Defense, Ministry of Education, Shanghai, China, [2]Shanghai Key Laboratory of Medical Bioprotection, Shanghai, China, [3]Department of Epidemiology, Second Military Medical University, Shanghai, China, [4]Department of Foreign Languages, International Exchange Center for Military Medicine, Second Military Medical University, Shanghai, China, [5]School of Medicine, Tongji University, Shanghai, China

Over three years' pandemic of 2019 novel coronavirus disease (COVID-19), multiple variants and novel subvariants have emerged successively, outcompeted earlier variants and become predominant. The sequential emergence of variants reflects the evolutionary process of mutation-selection-adaption of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Amino acid substitution/insertion/deletion in the spike protein causes altered viral antigenicity, transmissibility, and pathogenicity of SARS-CoV-2. Early in the pandemic, D614G mutation conferred virus with advantages over previous variants and increased transmissibility, and it also laid a conservative background for subsequent substantial mutations. The role of genomic recombination in the evolution of SARS-CoV-2 raised increasing concern with the occurrence of novel recombinants such as Deltacron, XBB.1.5, XBB.1.9.1, and XBB.1.16 in the late phase of pandemic. Co-circulation of different variants and co-infection in immunocompromised patients accelerate the emergence of recombinants. Surveillance for SARS-CoV-2 genomic variations, particularly spike protein mutation and recombination, is essential to identify ongoing changes in the viral genome and antigenic epitopes and thus leads to the development of new vaccine strategies and interventions.

KEYWORDS

SARS-CoV-2, evolution, Omicron, spike protein, amino acid substitution, recombination

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a sister clade of SARS-CoV (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses et al., 2020), has posed a global public health threat since its initial outbreak in December 2019 (Hu et al., 2020). On May 5, 2023, the World Health Organization (WHO) declared the end of the 2019 novel coronavirus disease (COVID-19) pandemic as a Public Health Emergency of International Concern. At that time, WHO reported a total of 765,222,932 cases and 6,921,614 deaths worldwide.[1] Consistent with other coronaviruses, the genome of SARS-CoV-2 is a single-stranded positive-sense RNA of approximately 30,000 nucleotides, with replication mediated by

1 https://covid19.who.int/

RNA-dependent RNA polymerase (RdRP) (Vkovski et al., 2020; Li et al., 2020b). The 5'-terminus of the SARS-CoV-2 genome contains two open reading frames (ORFs), while the 3'-terminus contains four major structural proteins coding-gene in the following order: spike protein, envelope protein, membrane protein, and nucleocapsid protein (Bai et al., 2021). Despite the presence of error-correction enzymes, which contribute to a relatively high replication fidelity compared to other RNA viruses, SARS-CoV-2 still undergoes significant mutations (Robson et al., 2020; Domingo et al., 2021; Perales, 2021). The nucleotide mutation rates of SARS-CoV-2 are estimated to be 6.677×10^{-4} and 8.066×10^{-4} substitutions per year for the whole genome and spike protein, respectively (Wang S. et al., 2021).

Amino acid mutations in the spike protein play a crucial role in the evolution of SARS-CoV-2. The spike protein, which forms a trimeric fusion protein on the surface of the coronavirus, exhibits a crown-like appearance and serves as an ideal target for inducing neutralizing antibodies and protective immunity (Kang et al., 2021; Tian et al., 2021). The spike protein is composed of S1 and S2 subunits, and the Receptor Binding Domain (RBD) in the spike interacts with the human receptor angiotensin-converting enzyme 2 (ACE2) receptor when activated to allow the virus to entry into cells (Conceicao et al., 2020; Hoffmann et al., 2020b; Zhang et al., 2021a). Mutations in the spike protein, particularly in the RBD, have led to alterations in spike-ACE2 recognition, resulting in viral immune escape and the failure of neutralizing antibodies (Magazine et al., 2022; Chen et al., 2023). Spike proteins are classified as open and closed forms according to the up and down conformations of the RBD, and mutations in the spike may change the RBD conformation (Walls et al., 2020; Wrapp et al., 2020). The D614G mutation, which represents the substitution of amino acid D (Asp) by G (Gly), is conservative across all major variants (Wassenaar et al., 2022) and predominant in the spike protein during the early stage of pandemic (Chang et al., 2020). The D614G mutation has been shown to enhance furin proteolysis capacity by 50 times (Gobeil et al., 2021). Notably, the Omicron variant harbors more than 60 substitutions, deletions, and insertions, of which 15 rare mutations are found in the spike (He et al., 2021; Ma et al., 2022b). The spike protein of Omicron predominantly adopts closed conformations (Calvaresi et al., 2023), potentially leading to the failure of nearly all anti-spike monoclonal antibodies (Focosi and Casadevall, 2022; Turelli et al., 2022).

In addition to point mutations in the spike protein, viral genomic recombination is common among coronaviruses (Yewdell, 2021), especially during the late pandemic phase when different variants co-circulate. According to the US Centers for Disease Control and Prevention (CDC), the most prevalent circulating strains in the US as of May 13, 2023, were XBB.1.5 (61.5%), XBB.1.9.1 (10.0%), and XBB.1.16 (9.4%) (Ma et al., 2023). The frequent occurrence of recombination makes it challenging to predict the effectiveness of vaccines targeting the spike protein, and recombination may confer altered transmissibility, virulence, and immune escape properties to the virus (Focosi and Maggi, 2022; Carabelli et al., 2023).

The evolution of SARS-CoV-2 within the population follows the mutation-selection-adaptation theory of Darwinian evolution (Goldman, 2021; Figure 1). In this context of hypermutation, both innate and adaptive host immune responses drive mutation selection (Thorne et al., 2021), as we have previously discussed (Shen et al., 2023). The virus evolves to adapt to external selection pressures, and antigenic drift occurs as mutations gradually accumulate, affecting the virus's immunogenicity (Bano et al., 2021; Shapira et al., 2023). Antigenic drift facilitates viral evasion from host immune response, particularly by affecting antibody neutralization, resulting in viral resistance to previous infection and vaccination (Zhang et al., 2022; Cao et al., 2022c; Planas et al., 2023; Qu et al., 2023). The evolutionary trend tends to lower the pathogenicity but increase the transmissibility of variants, resulting in long-term retention of virus in human hosts (Magiorkinis, 2023). In this review, we provide an overview of SARS-CoV-2, summarize the characteristic amino acid mutations in the spike protein, particularly in novel variants, discuss recent recombination events, and propose future perspectives to guide viral evolution and intervention strategies.

2. An overview of SARS-CoV-2

2.1. Nomenclature and timeline of SARS-CoV-2

Several nomenclatures have been introduced for SARS-CoV-2 according to genetic relatedness of the sequences, including GISAID,[2] Year-Letter (NextStrain) nomenclature,[3] and Phylogenetic Assignment of Named Global Outbreak LINeages (Pango lineage) (Rambaut et al., 2020). The GISAID nomenclature system is based on marker mutations within the eight high-level phylogenetic groups, from the early split of S and L, to the further evolution of L into V and G, and later G into GH, GR and GV, and more recently GR into GRY. The Year-Letter nomenclature consists of the year when the clade emerged and a capital letter starting with A for each year, including 19A, 19B, 20A, 20B, 20C, and 20I. The Pango lineage uses an alphabetical prefix and a numerical suffix to identify descendants[4] and contains phylogenetic, genetic, and epidemiological information. The first letter represents the lineage label of the variant, with the order from A to Z, then AA to AZ, BA to BZ, etc. The subsequent numbers separated by periods indicate the branches of lineages. When a branch has three more numeric suffixes, a new letter will be used as the lineage label in alphabetical order. For example, C.1 is the branch of B.1.1.1 (O'Toole et al., 2022). The recombinant variants are named in a uniform nomenclature beginning with "X."

To promote surveillance and research, WHO categorized SARS-CoV-2 variants as three specific classes: variants of concern (VOC), variants of interest (VOI), and variants under monitoring (VUMs).[5] VOCs are variants of high mutation and transmission rate. To date, Alpha, Beta, Gamma, Delta, and Omicron are known emerged VOCs and have become dominant in turn globally or regionally. The Alpha variant (B.1.1.7) was discovered in the UK in September 2020 (du Plessis et al., 2021; Galloway et al., 2021). It was proven to be highly transmissible and infectious, and became prevalent a few months later (Davies et al., 2021; Volz et al., 2021). The Beta variant (B.1.351) was

2 https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/

3 https://nextstrain.org/ncov

4 https://www.pango.network/the-pango-nomenclaturesystem/statement-of-nomenclature-rules/

5 https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/

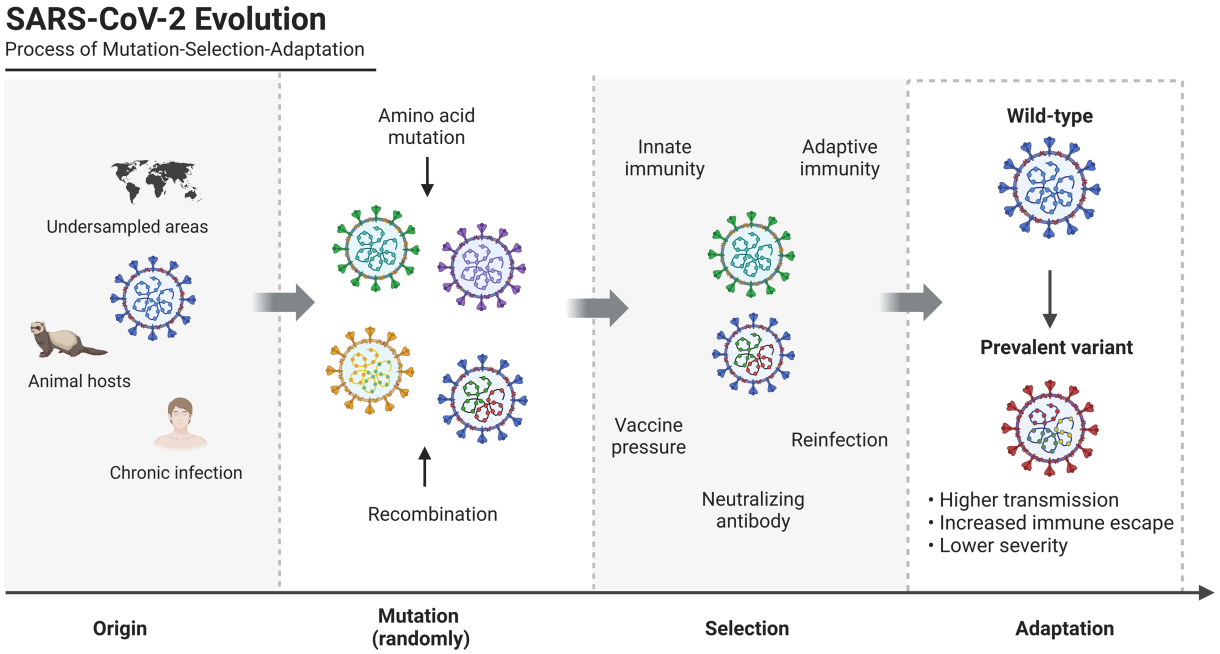


FIGURE 1
Process of mutation-selection-adaptation in SARS-CoV-2 evolution.

first reported in South Africa in October 2020 (Tegally et al., 2021), and the Gamma variant (P.1) was first identified in travelers from Brazil in January 2021 (Fujino et al., 2021). The Delta variant (B.1.617.2) was isolated in India (Mlcochova et al., 2021) and quickly became the most prevalent variant worldwide in June 2021 (Mahase, 2021). The Omicron variant (B.1.1.529/BA sublineages) was first discovered in Botswana, South Africa in November 2021, and outcompeted other VOCs rapidly upon its emergence (He et al., 2021). Five major sublineages of Omicron, BA.1, BA.2, BA.3, BA.4, and BA.5, have been identified so far (Tegally et al., 2022). Most recently, a series of novel Omicron subvariants have emerged, such as BA.2.75 (Saito et al., 2022), BF.7 (Scarpa et al., 2023a), Deltacron (Kreier, 2022), XE (Rahimi and Bezmin Abadi, 2022b), XF (Chakraborty et al., 2022), BQ.1 (Wang et al., 2022b), BQ.1.1 (Wang et al., 2022b), XBB (Imai et al., 2023), XBB.1 (Arora et al., 2023), XBB.1.5 (Tamura et al., 2022), XBB.1.16 (Harris, 2023), and they have raised increasing concern. The timeline of emergence of variants is illustrated in Figure 2A.

2.2. Entry pathways of SARS-CoV-2 and hypotheses for VOCs

Two described entry pathways of SARS-COV-2 through the cell membrane or through endosomes (Figure 2B) have been reviewed in detail previously (Shang et al., 2020; Hoffmann et al., 2020b; Rahbar Saadat et al., 2021; Jackson et al., 2021b; Lim, 2023). The two entry pathways differ because S2' cleavage occurs either at plasma membrane by the transmembrane protease serine protease 2 (TMPRSS2) [such as in the nasal epithelial cells, lungs, and bronchial branches where TMPRSS2 is highly co-expressed with ACE2 (Lukassen et al., 2020; Sungnak et al., 2020)] or within the cell by endolyosomal cathepsins such as Cathepsin L (Bestle et al., 2020; Shang et al., 2020). The proteolytic site between the S1 and S2 subunit of the spike protein, also known as furin cleavage site (FCS), is cleaved by a host protease furin (Lavie et al., 2022). This process of cleavage is essential to the entry pathway and membrane fusion (Bestle et al., 2020; Hossain et al., 2021; Johnson et al., 2021; Peacock et al., 2021; Lavie et al., 2022). Optimization of FCS has been shown to facilitate cell–cell fusion to improve the infectivity (Hoffmann et al., 2020a), increase the transmissibility (Peacock et al., 2021), and promote pathogenesis (Johnson et al., 2021).

Multiple hypotheses have been proposed to explain the origin of VOCs (Mallapaty, 2022), such as (1) circulation in geographically sequencing limited areas; (2) circulation within animal hosts then spillover to humans; and (3) evolution in immunosuppressed chronic infection hosts. In some regions, the limited capacity for genomic sequencing has resulted in a lack of testing for asymptomatic patients. It has been observed that asymptomatic carriers exhibit higher levels of antiviral immunity and lower levels of inflammation compared to symptomatic individuals (Yang et al., 2020b; Le Bert et al., 2021; Ma et al., 2022a). This immunological profile may create an environment conducive to viral evolution under immune pressure. There is evidence supporting the hypothesis of an animal host origin, with white-tailed deer (Hale et al., 2022; Marques et al., 2022) and farmed mink (Koopmans, 2021; Lu et al., 2021) identified as stable animal reservoirs for SARS-CoV-2. These variants have the potential to infect animals and accumulate mutations within animal reservoirs. Subsequently, the virus may undergo further evolution, giving rise to new subvariants that can then spillover to humans. The hypothesis of chronic infection in immunodeficient hosts is widely accepted in many scenarios. Chronic infection in such individuals is associated with ACE2 affinity, immune evasion, and optimization of viral packaging (Choi et al., 2020; Kemp et al., 2021; Harari et al., 2022;

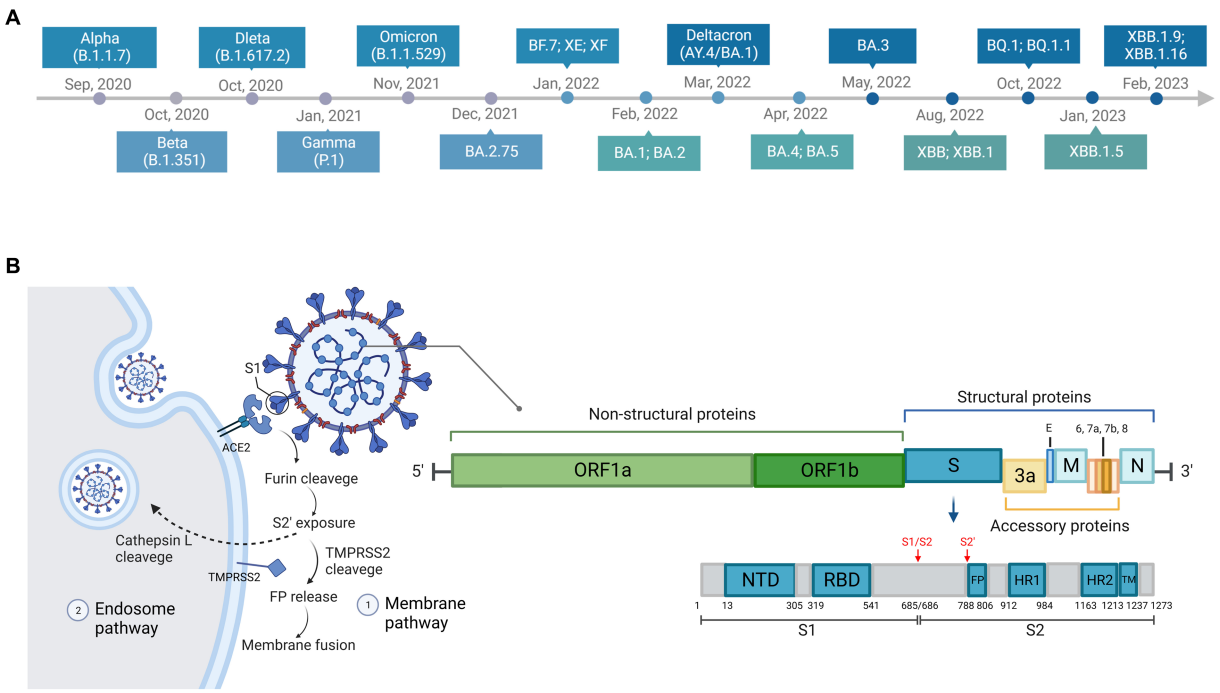


FIGURE 2
Timeline, structure, and entry pathways of SARS-CoV-2. **(A)** The chronological order of the emergence of major SARS-CoV-2 variants. **(B)** There are two pathways for SARS-CoV-2 entering cells: endosome pathway and membrane pathway. ACE2, angiotensin-converting enzyme 2; TMPRSS2, transmembrane protease serine protease 2; S1, subunit 1 of the spike protein; FP, fusion peptide, responsible for membrane fusion; S1/S2, furin cleavage site between S1 and S2 subunit of the spike protein; S2', another proteolytic site in the subunit 2 of the spike protein.

Wilkinson et al., 2022). This process drives the mutation profiles of the virus and enhances its fitness (Ghafari et al., 2022; Hill et al., 2022). Extensive immune escape has been observed in SARS-CoV-2 infections in immunocompromised hosts, such as patients with advanced HIV disease (Cele et al., 2022).

3. Spike protein mutations produce antigenic drift

Mutation profiles of the variants of concern (VOCs) exhibit certain overlapping patterns, while also assuming distinct roles in the process of viral evolution, thereby suggesting an underlying evolutionary resemblance among these variants. Notably, a common early substitution mutation, namely D614G, is shared by all five VOCs, which has been shown to significantly augment the binding affinity of the viral spike protein to the ACE2 receptor, consequently amplifying viral pathogenicity (Alkhatib et al., 2021; Wang P. et al., 2021; Zhang et al., 2021b; Venkatakrishnan et al., 2022). Moreover, the substitution P681H has been identified in Alpha (Lubinski et al., 2022), Gamma (Fujino et al., 2021), and Omicron (Tian et al., 2022), and has been demonstrated to enhance viral cell entry. Conversely, the substitution P681R, occurring at the same position, has been observed to augment the replication capacity and pathogenicity of the Delta variant (Mlcochova et al., 2021; Saito et al., 2021; Liu et al., 2022). These mutations accumulate in a stepwise manner, progressively modifying the antigenic epitope of the virus, ultimately leading to a transition from "genetic drift" to antigenic drift.

3.1. Spike mutations in current VOCs

For variant Alpha (B.1.1.7), of eight mutations in the spike protein, D614G, Del H69/V70 (Del H69/V70 represents amino acid deletion mutation in the site 69 and 70 of the spike protein), N501Y, and P681H are most meaningful (Wang P. et al., 2021). The D614G mutation has been found to confer a fitness advantage by promoting efficient replication in primary airway cells, thereby increasing virulence and transmission (Hou et al., 2020; Korber et al., 2020; Ozono et al., 2021; Zhou et al., 2021). It also leads to alterations in spike conformation and enhanced FCS cleavage (Zhang et al., 2020) and leads to alterations in spike conformation and enhanced FCS cleavage (Gobeil et al., 2021; Nguyen et al., 2021). However, it has also been observed that the D614G mutation renders the virus more susceptible to monoclonal antibodies by increasing epitope exposure, suggesting that it does not impede the effectiveness of vaccines (Weissman et al., 2020), indicating it does not impede vaccine effect (Hou et al., 2020; Weissman et al., 2020; Yurkovetskiy et al., 2020; Ozono et al., 2021). Del H69/V70 is associated with diagnostic test failure for probes targeting spike proteins, known as spike gene targeting failure (SGTF) (Bal et al., 2021). SGTF has been utilized as a reliable proxy for monitoring the prevalence of the B.1.1.7 variant (Bal et al., 2021; Borges et al., 2021; Kidd et al., 2021). N501Y has been shown to enhance the binding of the spike protein to human ACE2 receptors, potentially expanding the host range of SARS-CoV-2 (Starr et al., 2020; Chan et al., 2021; Zahradník et al., 2021; Wang et al., 2022c). P681H, which is located adjacent to the FCS, has been found to enhance the efficiency of FCS cleavage during virus entry into cells

and contributes to Alpha's resistance to type I interferons (Lubinski et al., 2022; Lista et al., 2022).

For Beta variant (B.1.351), the combination of E484K and N501Y mutations has a synergistic effect in enhancing the affinity of the spike protein for human ACE2 receptors (Starr et al., 2020; Zahradník et al., 2021). Mutations Del 242–244, K417N, E484K, and N501Y have been shown to confer significant resistance to infection or vaccine-induced neutralizing antibodies (Garcia-Beltran et al., 2021; Hu et al., 2021; Tao et al., 2021; Wibmer et al., 2021; Wang P. et al., 2021).

The Gamma variant (P.1) carries 12 mutations in the spike protein, including K417T, N501Y, and E484K (Faria et al., 2021). These three mutations collectively enhance the affinity of the spike protein for ACE2 receptors, thereby increasing the transmissibility of the Gamma variant. E484K is also associated with reduced neutralization by antibodies (Faria et al., 2021). E484K is also associated with reduced neutralization by antibodies (Cele et al., 2021; Greaney et al., 2021; Wibmer et al., 2021).

The Delta variant (B.1.617.2) harbors several mutations previously reported in other VOCs, including L452R, T478K, E484Q, D614G, and P681R in the spike protein (Liu et al., 2022). These mutations partly explain the rapid global spread of the Delta variant upon its emergence. The L452R mutation has been found to increase infectivity, modestly reduce susceptibility to neutralizing antibodies, and enhance viral fusogenicity, thereby promoting virus replication (Motozono et al., 2021). E484Q exhibits similar reduced sensitivity to vaccine-induced neutralizing antibodies as L452R, but lacks synergistic effects when taken together (Motozono et al., 2021). Similar to P681H in Alpha, P681R in Delta increases FCS cleavage, resulting in enhanced transmissibility (Mlcochova et al., 2021; Saito et al., 2021; Wibmer et al., 2021). Studies have revealed that spike of Delta is more stable and binds with higher affinity to ACE2 than the spike of the wild-type (Gomari et al., 2023).

As discussed above, the evolution of SARS-CoV-2 of pre-Omicron variants has primarily centered around recurrent mutations in key residues of the spike protein, including D614, N501, P681, K417, and E484. However, with the emergence of the Omicron variant and its sublineages, the landscape has undergone a significant shift. The Omicron variant harbors over 30 spike mutations, with 15 of them occurring in the RBD (Kumar et al., 2021). Figure 3 illustrates the mutation profiles of VOCs. In general, Omicron exhibits several distinctive characteristics compared to previous VOCs, including enhanced transmissibility, reduced antibody neutralization capacity (resulting in lower vaccine effectiveness), altered tissue tropism, relatively lower pathogenicity, and an increased likelihood of reinfection.

The higher transmissibility may attribute to the altered viral affinity to ACE2 receptor. Multiple experimental observations have demonstrated that the binding affinity between the RBD of the spike protein and ACE2 is significantly higher for Omicron compared to wildtypes (Kumar et al., 2021; Abeywardhana et al., 2022; Cui et al., 2022; Hong et al., 2022). The mutations T478K, Q493R, Q498R, and N501Y collectively contribute to the increased binding affinity through electrostatic effects (Kumar et al., 2021; Abeywardhana et al., 2022). However, another study revealed that Omicron exhibits comparable binding affinity to ACE2 when compared to the wild type SARS-CoV-2 and weaker binding affinity than the Delta variant (Wu et al., 2022). This discrepancy may stem from differences in the surface plasmon resonance methodologies employed in the studies, necessitating further research. The sublineages of Omicron display variations in their ACE2 affinity, with BA.2 exhibiting the highest affinity, followed by BA.3, BA.1, BA.2.75, and BA.5 (Abeywardhana et al., 2022). Furthermore, Omicron variants exhibit reduced sensitivity to neutralizing antibodies induced by triple-dose inactivated vaccines (Ren et al., 2023). Reports indicate that the neutralizing activity against Omicron variants is lost in 90% of immunization serum samples and 43% of convalescent serum samples (Zhang et al., 2022). In contrast to pre-Omicron variants, which primarily exploit TMPRSS2 for cell entry (Hoffmann et al., 2020b), Omicron variants have a propensity for entering nose and throat cells that are deficient in TMPRSS2 via the cathepsin-mediated endosomal pathway (Hui et al., 2022; Meng et al., 2022; Willett et al., 2022; Zhao et al., 2022). This shift in cell entry tropism from the membrane pathway to the endosomal pathway reduces the capacity of Omicron to fuse infected cells and form syncytia, resulting in a lower pathogenicity (Meng et al., 2022; Willett et al., 2022). The Omicron's propensity to infect upper respiratory tract restricts its clinical manifestation and lowers the disease severity. From a structural standpoint, compared with Delta, Omicron has an inconsistent distribution of electrostatic potential and a geometric reorganization in the FCS of the spike protein. This structural divergence contributes to Omicron's reduced fusogenicity and consequently lower pathogenicity (Fantini et al., 2022). Moreover, the Omicron variant possesses an enhanced capacity for immune evasion, leading to reinfection of individuals (Chavda et al., 2022; Xia et al., 2022). For pre-Omicron variants, infection-induced protective immunity has limited efficacy against BA.4 and BA.5, but it demonstrates a strong effect in preventing reinfection of BA.1 and BA.2 (Altarawneh et al., 2022).

Notably, the combinatorial mutations in the spike protein appear to have a synergistic effect on the characteristics of Omicron, further complicating its mutation profile. Preliminary findings suggest that certain mutations in Omicron form three distinct clusters, wherein the mutations seem to work in concert to compensate for the detrimental effects of any individual mutation (Martin et al., 2022). Two mutations, N501Y and Q498R, collectively increase the affinity of a variant for the ACE2 receptor by nearly 20-fold (Bate et al., 2022).

3.2. Spike mutations in novel subvariants

3.2.1. BA.2.75 (BM.1.1.1)

BA.2.75, a descendant from BA.2, was first detected in India and Singapore (Saito et al., 2022). Differing from BA.2, BA.2.75 carries 9 additional mutations in the spike protein (147E, W152R, F157L, I210V, G257S, D339H, G446S, N460K, and an R493Q reversion mutation) (Sheward et al., 2022; Kurhade et al., 2023; Qu et al., 2023). BA.2.75 exhibits enhanced resistance to neutralization compared to BA.2 but falls short of the BA.4/5 variant (Qu et al., 2022; Saito et al., 2022; Cao et al., 2022b; Wang L. et al., 2022). The G446S and N460K mutations are primarily responsible for the increased resistance of neutralizing antibodies against BA.2.75 (Qu et al., 2022; Wang et al., 2022a), while the R493Q mutation reduces neutralization resistance (Wang et al., 2022a). Furthermore, the spike protein of BA.2.75 demonstrates significantly higher affinity for ACE2 (Saito et al., 2022), and the N460K mutation, which enhances S processing, leads to

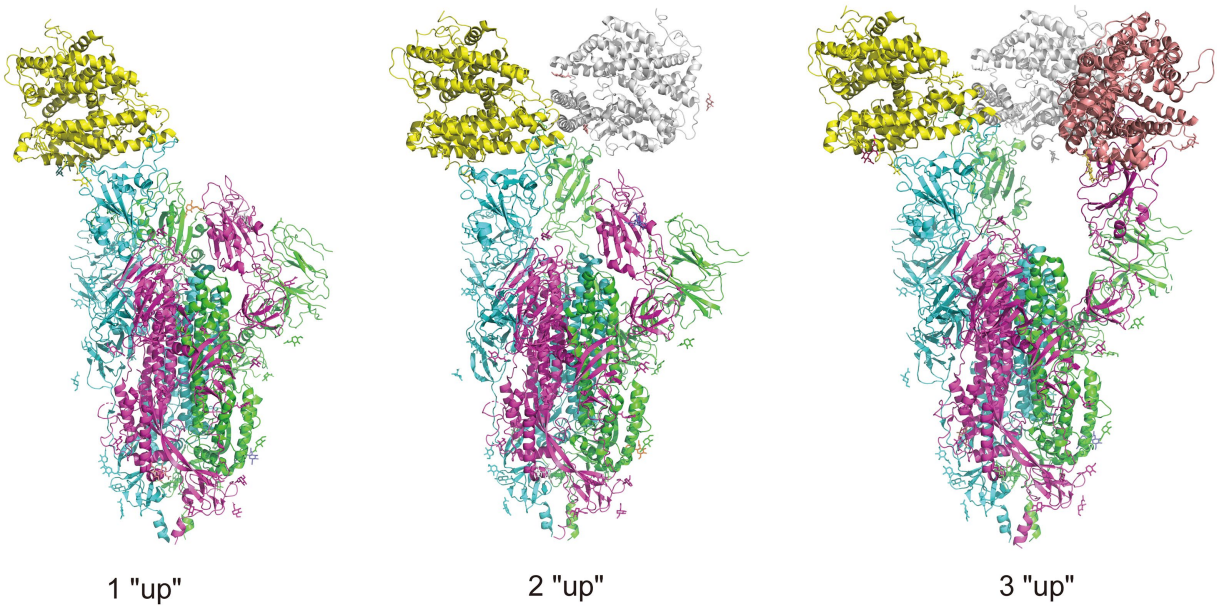


FIGURE 3
Illustration of RBD conformation of spike protein complexed with ACE2 receptors. There are two RBD conformations: "up" and "down," and when the RBD is in "up" conformation, the spike protein is open to the ACE2 receptor. The trimeric spike protein is indicated by chain in three colors, purple, green, and blue, and three ACE2 receptors are indicated in yellow, gray, and pink. The complexes are obtained from RCSB.org (7KNE, 7KNH, 7KNI for 1 "up," 2 "up," 3 "up," respectively).

increased cell–cell fusion of BA.2.75 compared to BA.2 (Qu et al., 2022).

3.2.2. BA.4.6

BA.4.6, a sublineage of BA.4, carries two additional mutations in the spike protein (R346T and N658S) and was initially identified in the US and UK (Hachmann et al., 2022). This subvariant exhibits a notable ability to evade neutralizing antibodies induced by infection or vaccination, with titers lower than those of BA.5 by a factor of 2 to 2.7 (Hachmann et al., 2022; Wang et al., 2022b; Planas et al., 2023).

3.2.3. BF.7

BF.7 variant (also known as BA.5.2.1.7) is a derivative of BA.5 and has gained attention since the beginning of 2022, particularly in Asia (Kelleni, 2023; Pan et al., 2023; Scarpa et al., 2023a). Compared to BA.5, BF.7 carries an additional R346T mutation in the RBD and shares an identical N-terminal domain (NTD) (Scarpa et al., 2023a). The R346T mutation has been associated with enhancing the virus's ability to evade neutralizing antibodies generated by vaccines or previous infection (Akif et al., 2023). However, R346T does not greatly increase the affinity of BF.7 to ACE2 (Scarpa et al., 2023a). Although enhanced resistance to neutralization exists (Qu et al., 2023), BF.7 appears to be less virulent, with a low evolutionary rate of 5.62×10^{-4} substitutions/sites/years compared to other Omicron subvariants (Scarpa et al., 2023a).

3.2.4. CH.1.1

CH.1.1, a descendant of BA.2.75, has rapidly emerged in the UK. Compared with BA.2.75, CH.1.1 owns additional 4 substitutions (R346T, K444T, L452R, and F486S) in the RBD of the spike protein (Uraki et al., 2023). CH.1.1 does not pose a significant threat to pandemic control. Antiviral drugs (remdesivir, molnupiravir, nirmatrelvir, and ensitrelvir) remain effective against CH.1.1, and an additional dose of bivalent mRNA vaccines may be beneficial in preventing CH.1.1 infection (Uraki et al., 2023).

3.2.5. BQ.1 and BQ.1.1

BQ.1 and BQ.1.1 have evolved from BA.5 (Wang et al., 2022b). Compared with the progenitor BA.5, BQ.1 carries additional K444T and N460K mutations in the spike protein, while BQ.1.1 has an additional R346T mutation (Wang et al., 2022b). Strong resistance to neutralization is observed in the BQ.1 and BQ.1.1 subvariants, largely driven by the N460K mutation (Kurhade et al., 2023; Qu et al., 2023).

3.2.6. XBB and XBB.1.5

XBB variant carries 9 additional changes in the RBD and 5 additional changes in the NTD compared to its progenitor BA.2 (Imai et al., 2023). The R346 position is a critical mutation site (harboring R346T/S/I) that leads to increased immune evasion by neutralizing antibodies (Cao et al., 2021). Similar to BQ.1 and BQ.1.1, the XBB lineage exhibits an exceptionally strong ability to evade antibodies (Arora et al., 2023). BQ and XBB subvariants have rendered all authorized antibodies ineffective, with titers against BQ and XBB significantly lower (Wang et al., 2022a; Chakraborty et al., 2023). A cohort study in Singapore revealed that protection against XBB reinfection was lower and weakened more rapidly compared to protection against BA.4 or BA.5 reinfection in previously vaccinated omicron-infected individuals (Tan et al., 2023), further indicating greater immune evasion in XBB.

XBB.1.5 has a substantial growth advantage over BQ.1.1 and XBB.1, becoming the predominant strain in the US by January 2023

(Tamura et al., 2022). XBB.1.5 is a recombinant of two descendants from BA.2, differing from XBB.1 by an additional F486P mutation in the spike protein (Tamura et al., 2022). Unlike the F486S mutation in XBB.1, which disrupts the local hydrophobic interaction of the spike with ACE2, F486P in XBB.1.5 restores this interaction (Yue et al., 2023). This mechanism enhances the affinity for ACE2 and suggests a higher growth advantage for XBB.1.5 compared to its progenitor XBB.1. F486P makes XBB.1.5 slightly less immune evasive but more infectious than its ancestor XBB.1, likely due to increased binding affinity to human ACE2 (Tamura et al., 2022; Mahase, 2023).

3.2.7. XBB.1.16

XBB.1.16 is another XBB sublineage harboring the F486P substitution, outcompeting other variants in India by the end of March 2023 (Looi, 2023; Varghese et al., 2023). Compared to XBB.1.5, XBB.1.16 carries two additional substitutions, E180V in the NTD and T478R in the RBD, in the spike protein (Yamasoba et al., 2023). XBB.1.16 exhibits a greater growth advantage compared to XBB.1 and XBB.1.5, but its potential for immune evasion is similar to XBB.1 and XBB.1.5 (Yamasoba et al., 2023). Notably, XBB.1.16 and XBB.1.5 demonstrate similar characteristics in terms of cell line tropism, cell entry efficiency, and neutralization evasion (Nehlmeier et al., 2023).

3.3. Spike mutations in RBD conformation

SARS-CoV-2 infection is partially controlled by the conformation of the spike protein RBD. The RBD located in the S1 subunit of the extracellular domain of the spike is responsible for interacting with ACE2 receptors, and has been shown an important molecular determinant of the COVID-19 pandemic (Shang et al., 2020). The RBD exists in two different conformations: up for receptor binding and down for immune evasion. Accordingly, the spikes are also in open and closed conformations. Compared with the closed-form spike protein, an open-form with an up RBD conformation leads to infection more rapidly (Yin et al., 2022), and binding with antibodies more easily (Berger and Schaffitzel, 2020; Yin et al., 2022). Figure 4 illustrates the different up or down conformations of spike protein complexed with ACE2 receptors. In the early phase of the pandemic, the D614G substitution adjacent to the NTD subdomain leads to a more open and thus receptor-accessible conformations of the spike compared with the wild-type (Benton et al., 2021; Gobeil et al., 2021; Mansbach et al., 2021; Zhang et al., 2021a). The D614G substitution confers the virus an adaptation advantage and higher transmissibility, facilitating the acquisition of further mutations and forming the variants of concern (Korber et al., 2020; Zhang et al., 2020; Plante et al., 2021). It is shown that the conformations of Alpha, Beta and Delta spikes are predominantly open and that the binding of ACE2 increases membrane fusion (Calvaresi et al., 2023). In contrast, substitution of the Omicron spikes results in a predominantly closed conformation that may allow them to evade antibodies (Calvaresi et al., 2023). Other studies show that the mutations in the RBD of Omicron may promote the conformation to change from "down" to "up" and thus increase engagement of ACE2 (Hossen et al., 2022; Ye et al., 2022). This may due to the mutations that reduce the protein–protein interaction affinity of RBD with its neighboring domains (Singh et al., 2022).

Glycosylation is another way to affect the RBD conformation and thus change the spike open state. The SARS-CoV-2 spike gene encodes 22 N-linked glycan sequons per protomer and the trimeric spike protein displays 66 N-linked glycosylation sites. Glycosylated spike has a higher barrier to opening and also energetically favors the down state over the up state (Pang et al., 2022). Inhibition of protein N-glycosylation is shown to block SARS-CoV-2 infection (Casas-Sanchez et al., 2021). The glycosylation sites also have the effect of facilitating immune evasion by shielding specific epitopes from antibody neutralization (Watanabe et al., 2019). It is observed that proximal glycosylation sites (N165, N234, and N343) shield the receptor binding sites on the SARS-CoV-2 spike, especially when the RBD is in the "down" conformation (Watanabe et al., 2020). Sztain et al. (2021) revealed that N-glycan at position N343 facilitates RBD opening, and plays a gating role in the spike protein open state. Although the spike surface is substantially shielded by N-glycans, it presents regions that are vulnerable to neutralizing antibodies such as in the RBM, NTD, and S2 subunit (Chi et al., 2020; Tortorici et al., 2020; Cerutti et al., 2021). Mutations in the spike may affect glycosylation. For example, P681H and P681R were found in Alpha and Delta, respectively, and they decreased O-glycosylation which potentially increases furin cleavage and may influence viral infectivity (Zhang et al., 2021c).

4. Recombinant mutations complement variants with new properties

Recombination, a frequently observed evolutionary mechanism in coronaviruses, plays a significant role in the genetic diversity and evolution of these viruses. For example, lineage 5 of Middle East respiratory syndrome coronavirus (MERS-CoV), which caused the MERS-CoV outbreak in South Korea and mass infections in Saudi Arabia in 2015, is putatively a recombinant virus of groups 3 and 5 of clade B, or lineages 3 and 4 (Wang et al., 2015; Sabir et al., 2016). The measurement of recombination versus *de novo* mutation (R/M) provides insights into the relative impact of these two variations (Patiño-Galindo et al., 2021). In SARS-CoV-2, the R/M ratio is 0.00264 (Turakhia et al., 2022), while in MERS, it is estimated to be 0.25–0.31 (Patiño-Galindo et al., 2021), indicating a low level of recombinant mutations in the early stage of the SARS-CoV-2 pandemic. However, as co-infections and mutation accumulation increase within the population, recombination is expected to play a more prominent role in generating functional genetic diversity (Kim et al., 2020).

4.1. Co-circulation of variants provides basis for recombination

Recombination occurs when genetically distinct SARS-CoV-2 variants co-infect the same host during co-circulation (Figure 5A). This process leads to the emergence of recombinant viruses with new properties, such as increased transmissibility or virulence (Li et al., 2020a). Recombination occurs frequently in the later phase of pandemic (Varabyou et al., 2021). Turakhia et al. (2022) developed a method called Recombination Inference using Phylogenetic PLacEmentS (RIPPLES) to detect recombination in pandemic-scale phylogenies. By analyzing a 1.6 million sample tree, they identified

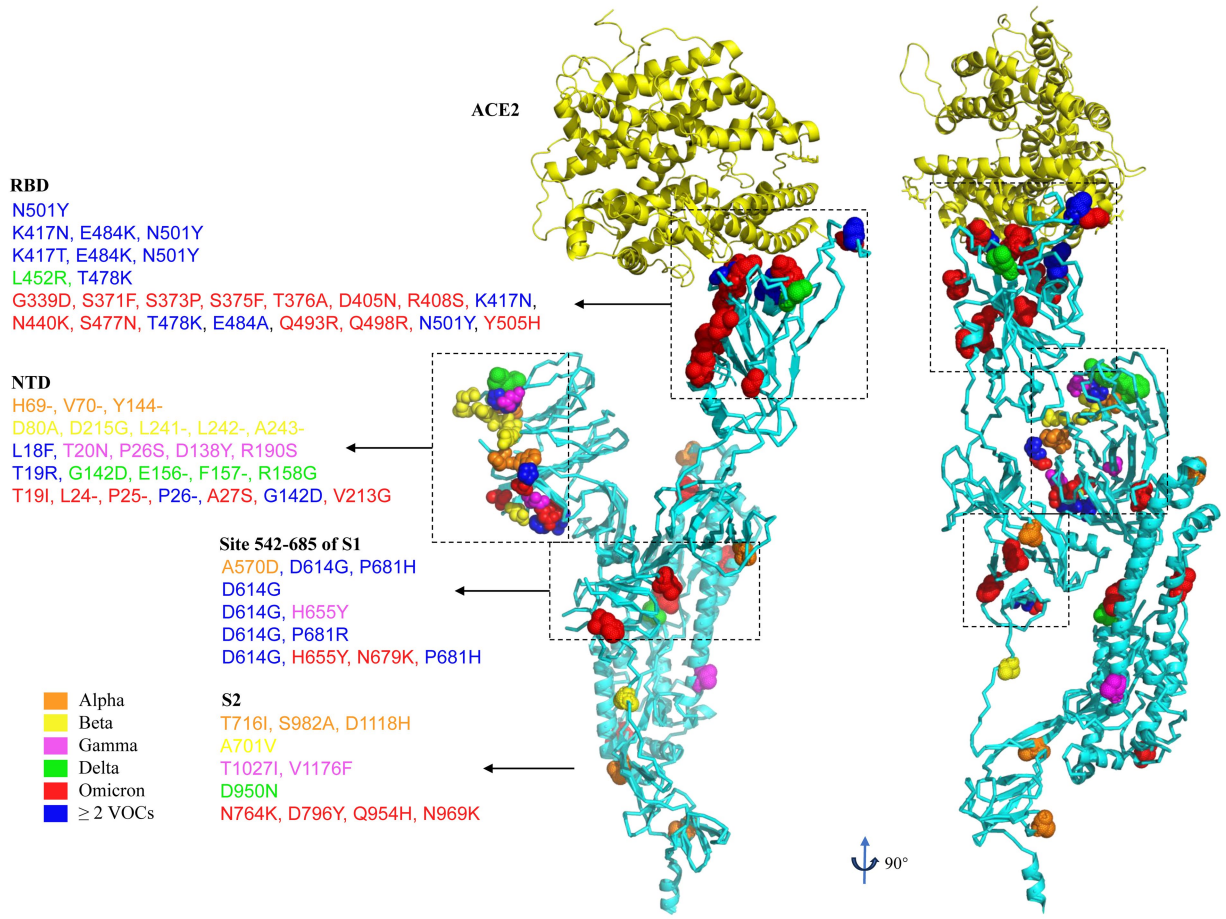


FIGURE 4
Representation of the SARS-CoV-2 spike protein, showing amino acid mutations in VOCs Alpha, Beta, Gamma, Delta and Omicron. Amino acid mutations are colored in orange, Alpha; yellow, Beta; purple, Gamma; green, Delta; red, Omicron; blue, ≥ 2 VOCs. The spike protein structure complexed with ACE2 receptor is obtained from RCSB.org (7KNE). The mutations of VOCs are based on the data from covariants (https://covariants.org, 20I for Alpha, 20H for Beta, 20 J for Gamma, 21A for Delta, and 21 L for Omicron).

589 recombination events, indicating that approximately 2.7% of sequenced SARS-CoV-2 genomes have detectable recombinant ancestry (Turakhia et al., 2022). The distribution of recombination breakpoints across the SARS-CoV-2 genome is not uniform, with a higher incidence toward the 3' end compared to the 5' end, consistent with previous analyses in other human coronaviruses (Patiño-Galindo et al., 2021; Müller et al., 2022). Recombination events often lead to genetic alterations near the breakpoints, and the specific breakpoints vary across the genome (Bolze et al., 2022). For example, a recombinant virus containing genetic material from the Alpha (B.1.1.7) and Epsilon (B.1.429) variants was detected in New York, and recombinant mutations were found in the spike, nucleocapsid, and ORF8 coding regions (Wertheim et al., 2022). In the US, there have been nine reported recombination events between the Delta (AY.119.2) and Omicron (BA.1.1) variants, with the breakpoint located between the NTD and RBD of the spike protein (Lacek et al., 2022a). These recombinants can produce hybridized spike proteins containing characteristic amino acids from both Delta and Omicron (Lacek et al., 2022a). The co-circulation of different variants highlights the importance of ongoing genomic surveillance, with particular attention to recombinants (Jackson et al., 2021a). Figure 5B illustrates different patterns of recombination.

4.2. Co-infection in immunocompromised population accelerates recombination

Co-infection is common in the later phase of the pandemic. For example, a 17-year-old Portuguese female was reported to be co-infected with two SARS-CoV-2 lineages belonging to distinct clades, differing by six variants (Pedro et al., 2021). Similar co-infection events have been observed, such as B.1.1.28 co-infecting with either B.1.1.248 or B.1.91 lineages (da Silva Francisco et al., 2021), and GH co-infecting with GR clades (Samoilov et al., 2021). In the US, out of 29,719 SARS-CoV-2 positive samples sequenced from November 2021 to February 2022, 20 co-infections were identified (Lacek et al., 2022b). In Brazil, nine co-infection events (0.61%) were identified in the investigated samples from May 2020 to April 2021, although this data is likely an underestimation due to sample limitations. Recombination has been found to occur more frequently in immunodeficient individuals at high risk of severe COVID-19

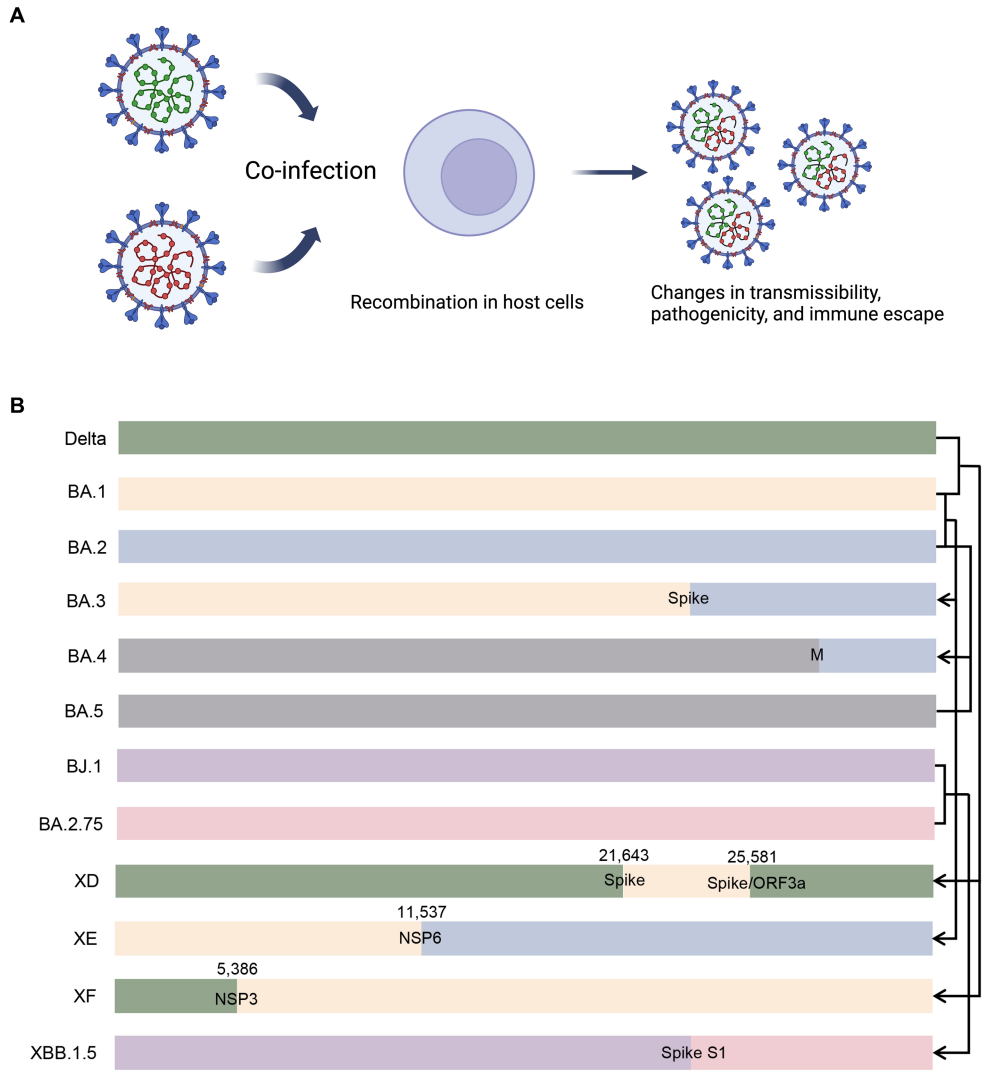


FIGURE 5
Illustration of recombination in co-infected cells and different recombination patterns. **(A)** When different variants co-infect an individual, there is possibility that recombinant variants emerge with altered properties. **(B)** BA.3 is putatively a recombinant of BA.1 and BA.2, and the breakpoint probably lies in the spike protein-coding gene. BA.4 is putatively a recombinant of BA.2 and BA.5, and the breakpoint probably lies in the M protein-coding gene. XD and XF are recombinants of Delta and BA.1, and the breakpoints lie in the spike protein-coding gene/ORF3a and NSP3 protein-coding gene, respectively. XE is a recombinant of BA.1 and BA.2, with breakpoint lying in the NSP6 protein-coding gene. XBB.1.5 is a recombinant od BJ.1 and BA.2.75, and the breakpoint probably lies in the S1 subunit of the spike protein-coding gene. M, membrane protein; ORF3a, open reading frame 3a; NSP, non-structural protein.

(Perez-Florido et al., 2023). Immunodeficient individuals are considered incubators for punctuated evolutionary events, possibly due to their vulnerability to chronic and co-infections (Rockett et al., 2022). For instance, a recombinant variant of B.1.160 and Alpha was isolated from a patient with lymphoma who was chronically infected for 14 months. The patient was initially infected with B.1.160, followed by concurrent Alpha infection, and eventually, the recombinant variant emerged (Burel et al., 2022).

4.3. Intra-variant recombination in omicron major subvariants

Recombination occurs in five major sublineages of Omicron. BA.1, a descendent lineage of B.1.1, shows distinctly different phylogenetic as compared with other VOCs or VOIs. It has caused the fourth epidemic wave in South Africa (Araf et al., 2022; Lino et al., 2022; Saxena et al., 2022; Tian et al., 2022). The spike gene sequencing reveals that the BA.1 subvariant shares nine common amino acid mutations with most VOCs in the spiked proteins (three more than BA.2) (Araf et al., 2022; Ou et al., 2022; Tian et al., 2022), suggesting that Omicron may be derived from the recombinant origin of these VOCs. Three more Alpha-associated mutations (Del 69, Del 70, and Del Y144) were found in BA.1 rather than in BA.2, for BA.1 is phylogenetically closer to Alpha than the other variants (Kumar et al., 2021; Ou et al., 2022). Reverse mutations were also found in some dominant mutations (frequency > 95%) in BA.1 (Ou et al., 2022). Taken together, these support the role of Alpha in Omicron evolution.

Along with BA.1, BA.2 and BA.3 were also isolated in South Africa (Zhou et al., 2022). BA.2 has caused increased global infection,

hospitalization, and mortality rate (Chen et al., 2022; Fonager et al., 2022; Rahimi and Bezmin Abadi, 2022a). BA.3 is likely a recombinant derivative of BA.1 and BA.2 due to BA.3 has similar genome in NTD region of the spike protein with BA.1 and BA.2 (Viana et al., 2022). A study revealed that BA.3 shared main mutations with BA.1 and BA.2, and BA.3 seemed to originate later (Wang C. et al., 2022), thus to some extent, corroborating the possibility of recombination.

BA.4 and BA.5 were afterwards identified as Omicron lineages in South Africa (Tegally et al., 2022). They were estimated to have originated in mid-December 2021 and early January 2022 (Viana et al., 2022). Their most recent common ancestor was estimated to have originated in mid-November 2021, coinciding with the emergence of BA.2 (Tegally et al., 2022). It deserves to note that BA.4 and BA.5 are close to BA.2 in genomes, and they both have similar spike proteins with BA.2 (Tegally et al., 2022). It is estimated that BA.4 and BA.5 are likely to evolve independently from the common ancestry of BA.2 subvariant (Wang C. et al., 2022). Compared with BA.2, BA.4 and BA.5 own extra mutations Del 69–70, L452R, F486V, and the wild-type amino acid at position Q493 (Ou et al., 2022). BA.4 and BA.5 share mutational profiles from 5'-UTR to envelope protein but differ distinctly from membrane protein to 3'-UTR (Tegally et al., 2022). This mutation pattern suggests that there exists a breakpoint within E and M, which is the possible evidence of recombinant event.

4.4. Inter-variant recombination between delta and omicron

Recombination events raised more concerns when Omicron quickly outcompeted Delta pandemic. Co-circulation of Delta and Omicron provided a grounded basis for recombinant variants. There is growing concern about the possibility that this recombination potential could eventually result in mutations that confer virus on enhanced transmissibility and immune escape properties.

On January 7, 2022, scientists detected a Delta and Omicron recombinant genome, and informally named it as "Deltacron" (Kreier, 2022). Nevertheless, it was later determined as a lab contamination (Kreier, 2022). On March 9, WHO declared the detection of such recombinants in different regions around the world and designated this Deltacron as a VUM (Farheen et al., 2022; Maulud et al., 2022). Generally, Deltacron is referred to as the AY.4/BA.1 recombinant, named XD, and consists of a full-length spike protein of Omicron and backbone of Delta (Mahase, 2022; Wang C. et al., 2022). According to Chinese Center for Disease Control and Prevention, of the 36 amino acid mutations found in the spike protein, 27 are present in BA.1 and 5 in AY.4, while 4 are present in both (Wang and Gao, 2022). Structural analysis of the Deltacron recombinant spike suggests its hybrid content leads to optimization of viral binding to the host cell membrane (Colson et al., 2022a,b). Consequently, this novel recombined virus causes increased disease transmission (Chakraborty et al., 2022; Hosch et al., 2022). The Deltacron recombinant also has the potential to escape neutralization by monoclonal antibody (Evans et al., 2022). Although Delta (AY.45) and BA.1 are sensitive to Sotrovimab neutralization, while an AY.45-BA.1 recombinant, with its breakpoint located adjacent to the Sotrovimab binding site, is resistant to its neutralization (Duerr et al., 2023). Deltacron shows higher transmissibility but lower clinical severity (Moisan et al., 2022). As recombination did not really emerge on a large scale and did not show its power until the appearance of Deltacron, the advent of Deltacron is regarded as a "gray rhino" event, rather than a "black swan" event.

Other than Deltacron (recombinant of AY.4 and BA.1, also known as XD), the UK Health security agency recognized two similar recombinants, XE and XF (Chakraborty et al., 2022). The XE recombinant contains genomic elements from Omicron BA.1 and BA.2 subvariants (Rahimi and Bezmin Abadi, 2022b). The breakpoint lies in the NSP6 protein-coding region of genome, with the 11,537 bp of the BA.1 and 11,537 bp of the BA.2 genomes before and after the break site (Chakraborty et al., 2022). XE appears to be roughly 10% more transmissible than its parent variant BA.2 (Basky and Vogel, 2022). The XF variant contains the genomes of NSP1 to NSP3 from the Delta variant; the breakpoint lies at site 5,386, and the rest genomes from Omicron BA.1 variant (Chakraborty et al., 2022).

XBB, nicknamed Gryphon, is the most recent recombinant. XBB is regarded as the first observed SARS-CoV-2 variant to increase its fitness through recombination rather than substitutions (Tamura et al., 2022). XBB derives from two BA.2 sublineages: BJ.1 (BA.2.10.1) and BM.1.1.1 (BA.2.75) (Arora et al., 2023; Scarpa et al., 2023b). XBB and its first descendant XBB.1 are both evolutionarily close to BA.2 genomes (Scarpa et al., 2023b), suggesting BA.2 acts as their progenitor. The breakpoint lies between position 22,901 and 22,939, a position in the middle of RBD (Scarpa et al., 2023b). The mutation profiles possibly altogether contribute to the greater immune invasion capabilities of XBB than do those of the earlier Omicron variants BA.2 (Imai et al., 2023). The pathogenicity of XBB.1 is comparable to or even lower than that of BA.2.75 (Tamura et al., 2022). Though XBB subvariants exhibit enhanced fusogenicity and substantial immune evasion in elderly population, but the fusion inhibitors EK1 and EK1C4 can potently block either XBB or XBB.1.5 spike protein mediated fusion and viral entry (Xia et al., 2023a).

4.5. Overall characteristics of emerging recombinants

As a whole, the novel recombinant subvariants demonstrate a higher transmission rate and relatively greater resistance to antibodies compared to earlier variants (Wang et al., 2022a; Brandolini et al., 2023; Faraone et al., 2023). In January 2023, there was a rapid increase in the prevalence of XBB.1.5 in the United States (Callaway, 2023). According to the World Health Organization (WHO), XBB.1.5 accounted for 23-86% of circulating variants throughout the country (XBB.1.5 Updated Risk Assessment, 24 February 2023).[6] However, these recombinant variants do not significantly increase the severity of the disease or cause clinical exacerbation (Karyakarte et al., 2023). XBB.1.5 does not carry mutations associated with potential changes in pathogenicity, such as P681R (Mlcochova et al., 2021; Saito et al., 2021). It is important to note that most vaccines are developed based on the spike protein, and the emergence of recombinant variants may pose a risk of vaccine failure (Tamura et al., 2022). Therefore, it is crucial to consider potential new subvariants in the development of novel strategic vaccines.

6 https://www.who.int/

5. Outlook for SARS-CoV-2 evolution and interventional strategies

Various factors drive the viral evolution (Moelling, 2021), including RNA polymerase exchanging accuracy for efficiency (Yewdell, 2021), the selective pressures exerted by host immune system (Milne et al., 2021; Thorne et al., 2021), chronic infection in other species then spillover to human (Lu et al., 2021; Hale et al., 2022; Marques et al., 2022), and prolonged co-infection in immunodeficient hosts (Ou et al., 2022; Rockett et al., 2022). These factors contribute to the mutation-selection-evolution process of SRAS-CoV-2 evolution. Continuous evolution of SARS-CoV-2 has led to rapid and simultaneous emergence of multiple variants that exhibit a growth advantage over previously circulating variants (Wolf et al., 2022). During the evolution of SARS-CoV-2, the spike gene is the only gene that undergo the strong positive selection, while other genes show only weak or temporary positive selection (Lu et al., 2023) Thus, spike mutations contribute highly in its evolution. The mutational process is dynamic, and the mutation spectrum of SARS-CoV-2 may tend to be more similar to that of other animal Sarbecoviruses (Bloom et al., 2022). Here we propose several interventional strategies.

1. Genomic surveillance of SARS-CoV-2, specifically in the spike gene and genomic recombination, is of utmost importance in recognizing its evolutionary trend. Efforts have been made to promote the genomic monitoring. Dadonaite et al. (2023) developed a novel deep mutational scanning (DMS) platform for mapping the effects of spike protein mutations on immune evasion and viral infectivity (Xia et al., 2023b). Saldivar-Espinoza et al. (2023) developed a SARS-CoV-2 Mutation Portal which provides access to a database of SARS-CoV-2 mutations. Sathyaseelan et al. (2023) developed a CoVe-tracker (SARS-CoV-2 evolution tracker)[7] for quick surveillance of newly emerging mutations/variants/lineages to facilitate the understanding of viral evolution, transmission, and disease epidemiology. Huang et al. (2023) developed a genomic surveillance framework and a dynamic community-based variant dictionary tree, which enables early detection and continuous investigation of SARS-CoV-2 variants. Outbreak. info is a platform for scalable and dynamic surveillance of SARS-CoV-2 variants and mutations, and it relies on shared virus sequences from the GISAID Initiative (Gangavarapu et al., 2023; Tsueng et al., 2023).
2. As the recently expanding Omicron subvariants are capable of immune evasion from most of the existing neutralizing antibodies, it is imperative to explore broad-spectrum antivirals to combat the emerging variants. Resistance to monoclonal antibody neutralization is dominated by the action of epitope single amino acid substitutions in the spike protein (Cox et al., 2022). Currently, most therapeutic neutralizing antibodies and promising vaccine candidates are designed to target the RBD or use RBD as the sole antigen (Shi et al., 2020; Yang et al., 2020a, 2021; Dai et al., 2022; Han et al., 2022). A novel group of neutralizing antibodies and vaccines targeting S2 subunit of the spike [such as fusion peptide (FP), heptad repeats 1 and 2 (HR1-HR2), and stem helix (SH)] may become the next generation of therapeutic strategies. For example, COV44-62 and COV44-79 were identified as anti-FP antibodies and showed considerable neutralizing capacity (Dacon et al., 2022).
3. Strategies should be implemented to prevent long-term SARS-CoV-2 infection and to limit the spread of emerging, neutralization-resistant variants in immunocompromised patients (Gonzalez-Reiche et al., 2023). It is found that the evolutionary rate of SARS-CoV-2 in chronic infection individual is 2-fold higher than that around the globe (Chaguza et al., 2023). This persistent intrahost evolution may accelerate antigenic alteration and lead to the emergence of genetically distinct subvariants (Smith and Ashby, 2022; Ahmadi et al., 2023; Chaguza et al., 2023). Bendall et al. (2023) observed a tight transmission bottleneck that would limit the development of highly mutated VOCs in the transmission chain of acutely infected individuals, further suggesting that selection for long-term infection in immunocompromised patients may drive SARS-CoV-2 VOC evolution (Braun et al., 2021; Wilkinson et al., 2022). Surveillance by sequencing is recommended for (i) patients carried with SARS-CoV-2, (ii) patients suspected of reinfection, and (iii) patients who are immunocompromised (Landis et al., 2023).
4. Vaccination in large population acts as a valuable measure in decreasing the mortality. However, vaccination alone cannot slow the pace of viral evolution for immune evasion and therefore, vaccine protection against severe and fatal outcomes for COVID-19 patients may not be assured (Van Egeren et al., 2023). Current herd immunity and BA.5 vaccine boosters may not efficiently prevent the infection of Omicron convergent variants (Cao et al., 2022a). However, these may result from the decreased pathogenicity of SARS-CoV-2 via inducing the mutations. The vaccination against SARS-CoV-2 still efficiently decrease the case fatality rate (Wang C. et al., 2022).

6. Summary and conclusion

In the process of SARS-CoV-2 evolution, external and internal pressures drive the selection of randomly occurring mutations, with the retention of favorable mutations leading to adaptation. SARS-CoV-2 exhibits a trajectory of evolution characterized by increased transmissibility, reduced virulence, and enhanced immune escape, enabling its long-term persistence within the population. The mutation patterns observed in pre-Omicron variants primarily manifest at recurrent amino acid sites within the spike protein, affecting the RBD conformation and glycosylation sites, consequently altering antigenicity. However, the emergence of the Omicron introduced a multitude of novel mutations, resulting in a substantial increase in transmissibility and immune evasion. Remarkably, the severity and clinical manifestation of patients did not escalate further, mainly for Omicron's tropism for the upper respiratory tract. These changes observed in Omicron are attributed to the ongoing viral evolution. The appearance of the recombinant variant XBB and its subsequent descendants since August 2022 likely stems from the co-circulation of multiple variants and co-infection in the immunocompromised patients during the later stage of the pandemic. Although novel recombinant variants such as XBB.1.5 and XBB.1.16 demonstrate a considerable transmission advantage and outcompete the predecessors, they do not exhibit a significant increase in disease severity and display

7 https://project.iith.ac.in/cove-tracker/

relatively moderate antibody escape. Although SARS-CoV-2 is no longer regarded as a Public Health Emergency of International Concern, its evolution persists. We strongly recommend for enhanced surveillance of the viral genome, particularly in immunocompromised patients, the development of therapeutics targeting domains beyond the RBD, and the promotion of widespread vaccination.

Author contributions

GC and LF: conceptualization. LF, JX, YZ, JF, and JS: data collection. LF, JX, and GC: writing—original draft preparation. LF, JX, GC, YZ, JF, WL, and JS: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (82041022) and Shanghai Commission of Science and Technology (20JC1410200 and 20431900404).

Acknowledgments

The authors acknowledge the use of Biorender.com to create Figure 1, 2, and 5.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abeywardhana, S., Premathilaka, M., Bandaranayake, U., Perera, D., and Peiris, L. D. C. (2022). *In silico* study of SARS-CoV-2 spike protein RBD and human ACE-2 affinity dynamics across variants and omicron subvariants. *J. Med. Virol.* 95:28406. doi: 10.1002/jmv.28406

Ahmadi, A. S., Zadheidar, S., Sadeghi, K., Nejati, A., Salimi, V., Hajiabdolbaghi, M., et al. (2023). SARS-CoV-2 intrahost evolution in immunocompromised patients in comparison with immunocompetent populations after treatment. *J. Med. Virol.* 95:e28877. doi: 10.1002/jmv.28877

Akif, A., Bhuiyan, M. A., and Islam, M. R. (2023). SARS-COV-2 omicron subvariant BF.7 is again triggering the Covid fear: what we need to know and what we should do? *J. Med. Virol.* 95:28551. doi: 10.1002/jmv.28551

Alkhatib, M., Svicher, V., Salpini, R., Ambrosio, F. A., Bellocchi, M. C., Carioti, L., et al. (2021). SARS-CoV-2 variants and their relevant mutational profiles: update summer 2021. *Microbiol. Spectr.* 9:e0109621. doi: 10.1128/spectrum.01096-21, 36198139

Altarawneh, H. N., Chemaitelly, H., Ayoub, H. H., Hasan, M. R., Coyle, P., Yassine, H. M., et al. (2022). Protective Effect of Previous SARS-CoV-2 Infection against Omicron BA.4 and BA.5 Subvariants. *N Engl J Med* 387, 1620–1622. doi: 10.1056/nejmc2209306

Araf, Y., Akter, F., Tang, Y., Fatemi, R., Parvez, M. D. S. A., Zheng, C., et al. (2022). Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* 94, 1825–1832. doi: 10.1002/jmv.27588

Arora, P., Cossmann, A., Schulz, S. R., Ramos, G. M., Stankov, M. V., Jäck, H.-M., et al. (2023). Neutralisation sensitivity of the SARS-CoV-2 XBB.1 lineage. *Lancet Infect. Dis.* 23, 147–148. doi: 10.1016/S1473-3099(22)00831-3

Bai, C., Zhong, Q., and Gao, G. F. (2021). Overview of SARS-CoV-2 genome-encoded proteins. *Sci. China Life Sci.* 65, 280–294. doi: 10.1007/s11427-021-1964-4

Bal, A., Destras, G., Gaymard, A., Stefic, K., Marlet, J., Eymieux, S., et al. (2021). Two-step strategy for the identification of SARS-CoV-2 variant of concern 202012/01 and other variants with spike deletion H69–V70, France, august to December 2020. *Euro Surveill.* 26. doi: 10.2807/1560-7917.ES.2021.26.3.2100008

Bano, I., Sharif, M., and Alam, S. (2021). Genetic drift in the genome of SARS COV-2 and its global health concern. *J. Med. Virol.* 94, 88–98. doi: 10.1002/jmv.27337

Basky, G., and Vogel, L. (2022). XE, XD & XF: what to know about the omicron hybrid variants. *CMAJ* 194, E654–E655. doi: 10.1503/cmaj.1095998

Bate, N., Savva, C. G., Moody, P. C. E., Brown, E. A., Evans, S. E., Ball, J. K., et al. (2022). *In vitro* evolution predicts emerging SARS-CoV-2 mutations with high affinity for ACE2 and cross-species binding. *PLoS Pathog.* 18:e1010733. doi: 10.1371/journal.ppat.1010733

Bendall, E. E., Callear, A. P., Getz, A., Goforth, K., Edwards, D., Monto, A. S., et al. (2023). Rapid transmission and tight bottlenecks constrain the evolution of highly transmissible SARS-CoV-2 variants. *Nat. Commun.* 14:272. doi: 10.1038/s41467-023-36001-5

Benton, D. J., Wrobel, A. G., Roustan, C., Borg, A., Xu, P., Martin, S. R., et al. (2021). The effect of the D614G substitution on the structure of the spike glycoprotein of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 118. doi: 10.1073/pnas.2022586118

Berger, I., and Schaffitzel, C. (2020). The SARS-CoV-2 spike protein: balancing stability and infectivity. *Cell Res.* 30, 1059–1060. doi: 10.1038/s41422-020-00430-4

Bestle, D., Heindl, M. R., Limburg, H., van Lam van, T., Pilgram, O., Moulton, H., et al. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci. Alliance* 3:e202000786. doi: 10.26508/lsa.202000786

Bloom, J. D., Beichman, A. C., Neher, R. A., and Harris, K. (2022). Evolution of the SARS-CoV-2 mutational spectrum. *bioRxiv*. doi: 10.1101/2022.11.19.517207

Bolze, A., Basler, T., White, S., Dei Rossi, A., Wyman, D., Dai, H., et al. (2022). Evidence for SARS-CoV-2 Delta and omicron co-infections and recombination. *Med* 3, 848–859.e4. doi: 10.1016/j.medj.2022.10.002

Borges, V., Sousa, C., Menezes, L., Gonçalves, A. M., Picão, M., Almeida, J. P., et al. (2021). Tracking SARS-CoV-2 lineage B.1.1.7 dissemination: insights from nationwide spike gene target failure (SGTF) and spike gene late detection (SGTL) data, Portugal, week 49 2020 to week 3 2021. *Euro Surveill.* 26. doi: 10.2807/1560-7917.ES.2021.26.10.2100130

Brandolini, M., Gatti, G., Grumiro, L., Zannoli, S., Arfilli, V., Cricca, M., et al. (2023). Omicron sub-lineage BA.5 and recombinant XBB evasion from antibody neutralisation in BNT162b2 vaccine recipients. *Microorganisms* 11:191. doi: 10.3390/microorganisms11010191

Braun, K., Moreno, G., Wagner, C., Accola, M. A., Rehrauer, W. M., Baker, D., et al. (2021). Limited within-host diversity and tight transmission bottlenecks limit SARS-CoV-2 evolution in acutely infected individuals. *bioRxiv*. doi: 10.1101/2021.04.30.440988

Burel, E., Colson, P., Lagier, J.-C., Levasseur, A., Bedotto, M., Lavrard-Meyer, P., et al. (2022). Sequential appearance and isolation of a SARS-CoV-2 recombinant between two major SARS-CoV-2 variants in a chronically infected immunocompromised patient. *Viruses* 14:1266. doi: 10.3390/v14061266

Callaway, E. (2023). Coronavirus variant XBB.1.5 rises in the United States — is it a global threat? *Nature* 613, 222–223. doi: 10.1038/d41586-023-00014-3

Calvaresi, V., Wrobel, A. G., Toporowska, J., Hammerschmid, D., Doores, K. J., Bradshaw, R. T., et al. (2023). Structural dynamics in the evolution of SARS-CoV-2 spike glycoprotein. *Nat. Commun.* 14:1421. doi: 10.1038/s41467-023-36745-0

Cao, Y., Jian, F., Wang, J., Yu, Y., Song, W., Yisimayi, A., et al. (2022a). Imprinted SARS-CoV-2 humoral immunity induces convergent omicron RBD evolution. *Nature* 614, 521–529. doi: 10.1038/s41586-022-05644-7

Cao, Y., Song, W., Wang, L., Liu, P., Yue, C., Jian, F., et al. (2022b). Characterization of the enhanced infectivity and antibody evasion of omicron BA.2.75. *Cell Host Microbe* 30, 1527–1539.e5. doi: 10.1016/j.chom.2022.09.018

Cao, Y., Wang, J., Jian, F., Xiao, T., Song, W., Yisimayi, A., et al. (2021). Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* 602, 657–663. doi: 10.1038/s41586-021-04385-3

Cao, Y., Yisimayi, A., Jian, F., Song, W., Xiao, T., Wang, L., et al. (2022c). BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by omicron infection. *Nature* 608, 593–602. doi: 10.1038/s41586-022-04980-y

Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., COVID-19 Genomics UK Consortium, et al. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* 21, 162–177. doi: 10.1038/s41579-022-00841-7

Casas-Sanchez, A., Romero-Ramirez, A., Hargreaves, E., Ellis, C. C., Grajeda, B. I., Estevao, I. L., et al. (2021). Inhibition of protein N-glycosylation blocks SARS-CoV-2 infection. *mBio* 13:21. doi: 10.1128/mbio.03718-21

Cele, S., Gazy, I., Jackson, L., Hwa, S. H., Tegally, H., Lustig, G., et al. (2021). Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* 593, 142–146. doi: 10.1038/s41586-021-03471-w

Cele, S., Karim, F., Lustig, G., San, J. E., Hermanus, T., Tegally, H., et al. (2022). SARS-CoV-2 prolonged infection during advanced HIV disease evolves extensive immune escape. *Cell Host Microbe* 30, 154–162.e5. doi: 10.1016/j.chom.2022.01.005

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., et al. (2021). Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* 29, 819–833.e7. doi: 10.1016/j.chom.2021.03.005

Chaguza, C., Hahn, A. M., Petrone, M. E., Zhou, S., Ferguson, D., Breban, M. I., et al. (2023). Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. *Cell. Rep. Med.* 4:100943. doi: 10.1016/j.xcrm.2023.100943

Chakraborty, C., Bhattacharya, M., Chopra, H., Islam, M. A., Saikumar, G., and Dhama, K. (2023). The SARS-CoV-2 omicron recombinant subvariants XBB, XBB.1, and XBB.1.5 are expanding rapidly with unique mutations, antibody evasion, and immune escape properties – an alarming global threat of a surge in COVID-19 cases again? *Int. J. Surg.* 109, 1041–1043. doi: 10.1097/JS9.0000000000000246

Chakraborty, C., Bhattacharya, M., Sharma, A. R., and Dhama, K. (2022). Recombinant SARS-CoV-2 variants XD, XE, and XF: the emergence of recombinant variants requires an urgent call for research – correspondence. *Int. J. Surg.* 102:106670. doi: 10.1016/j.ijsu.2022.106670

Chang, T.-J., Yang, D.-M., Wang, M.-L., Liang, K.-H., Tsai, P.-H., Chiou, S.-H., et al. (2020). Genomic analysis and comparative multiple sequences of SARS-CoV2. *J. Chin. Med. Assoc.* 83, 537–543. doi: 10.1097/jcma.0000000000000335

Chan, K. K., Tan, T. J. C., Narayanan, K. K., and Procko, E. (2021). An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci. Adv.* 7:eabf1738. doi: 10.1126/sciadv.abf1738

Chavda, V. P., Patel, A. B., and Vaghasiya, D. D. (2022). SARS-CoV-2 variants and vulnerability at the global level. *J. Med. Virol.* 94, 2986–3005. doi: 10.1002/jmv.27717

Chen, L.-L., Abdullah, S. M. U., Chan, W.-M., Chan, B. P.-C., Ip, J. D., Chu, A. W.-H., et al. (2022). Contribution of low population immunity to the severe omicron BA.2 outbreak in Hong Kong. *Nat. Commun.* 13:3618. doi: 10.1038/s41467-022-31395-0

Chen, Z., Li, J., Zheng, J., Jin, Y., Zhang, Y., Tang, F., et al. (2023). Emerging omicron subvariants evade neutralizing immunity elicited by vaccine or BA.1/BA.2 infection. *J. Med. Virol.* 95. doi: 10.1002/jmv.28539

Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., et al. (2020). A neutralizing human antibody binds to the N-terminal domain of the spike protein of SARS-CoV-2. *Science* 369, 650–655. doi: 10.1126/science.abc6952

Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., et al. (2020). Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* 383, 2291–2293. doi: 10.1056/NEJMc2031364

Colson, P., Delerce, J., Beye, M., Levasseur, A., Boschi, C., Houhamdi, L., et al. (2022a). First cases of infection with the 21L/BA.2 omicron variant in Marseille, France. *J. Med. Virol.* 94, 3421–3430. doi: 10.1002/jmv.27695

Colson, P., Fournier, P., Delerce, J., Million, M., Bedotto, M., Houhamdi, L., et al. (2022b). Culture and identification of a "Deltamicron" SARS-CoV-2 in a three cases cluster in southern France. *J. Med. Virol.* 94, 3739–3749. doi: 10.1002/jmv.27789

Conceicao, C., Thakur, N., Human, S., Kelly, J. T., Logan, L., Bialy, D., et al. (2020). The SARS-CoV-2 spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol.* 18:e3001016. doi: 10.1371/journal.pbio.3001016

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat. Microbiol.* 5, 536–544. doi: 10.1038/s41564-020-0695-z

Cox, M., Peacock, T. P., Harvey, W. T., Hughes, J., Wright, D. W., Willett, B. J., et al. (2022). SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro studies. *Nat. Rev. Microbiol.* 21, 112–124. doi: 10.1038/s41579-022-00809-7

Cui, Z., Liu, P., Wang, N., Wang, L., Fan, K., Zhu, Q., et al. (2022). Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 omicron. *Cells* 185, 860–871.e13. doi: 10.1016/j.cell.2022.01.019

Dacon, C., Tucker, C., Peng, L., Lee, C.-C. D., Lin, T.-H., Yuan, M., et al. (2022). Broadly neutralizing antibodies target the coronavirus fusion peptide. *Science* 377, 728–735. doi: 10.1126/science.abq3773

Dadonaite, B., Crawford, K. H. D., Radford, C. E., Farrell, A. G., Yu, T. C., Hannon, W. W., et al. (2023). A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cells* 186, 1263–1278.e20. doi: 10.1016/j.cell.2023.02.001

Dai, L., Gao, L., Tao, L., Hadinegoro, S. R., Erkin, M., Ying, Z., et al. (2022). Efficacy and safety of the RBD-dimer–based COVID-19 vaccine ZF2001 in adults. *N. Engl. J. Med.* 386, 2097–2111. doi: 10.1056/NEJMoa2202261

da Silva Francisco, R., Benites, L. F., Lamarca, A. P., de Almeida, L. G. P., Hansen, A. W., Gularte, J. S., et al. (2021). Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. *medRxiv* 15:198345. doi: 10.1101/2021.01.21.21249764

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372. doi: 10.1126/science.abg3055

Domingo, E., García-Crespo, C., Lobo-Vega, R., and Perales, C. (2021). Mutation rates, mutation frequencies, and proofreading-repair activities in RNA virus genetics. *Viruses* 13:1882. doi: 10.3390/v13091882

Duerr, R., Zhou, H., Tada, T., Dimartino, D., Marier, C., Zappile, P., et al. (2023). Delta-omicron recombinant escapes therapeutic antibody neutralization. *iScience* 26:106075. doi: 10.1016/j.isci.2023.106075

du Plessis, L., McCrone, J. T., Zarebski, A. E., Hill, V., Ruis, C., Gutierrez, B., et al. (2021). Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* 371, 708–712. doi: 10.1126/science.abf2946

Evans, J. P., Qu, P., Zeng, C., Zheng, Y.-M., Carlin, C., Bednash, J. S., et al. (2022). Neutralization of the SARS-CoV-2 Deltacron and BA.3 variants. *N. Engl. J. Med.* 386, 2340–2342. doi: 10.1056/NEJMc2205019

Fantini, J., Yahi, N., Colson, P., Chahinian, H., La Scola, B., and Raoult, D. (2022). The puzzling mutational landscape of the SARS-2-variant omicron. *J. Med. Virol.* 94, 2019–2025. doi: 10.1002/jmv.27577

Faraone, J. N., Qu, P., Evans, J. P., Zheng, Y.-M., Carlin, C., Anghelina, M., et al. (2023). Neutralization escape of omicron XBB, BR.2, and BA.2.3.20 subvariants. *Cell. Rep. Med.* 4:101049. doi: 10.1016/j.xcrm.2023.101049

Farheen, S., Araf, Y., Tang, Y., and Zheng, C. (2022). The Deltacron conundrum: its origin and potential health risks. *J. Med. Virol.* 94, 5096–5102. doi: 10.1002/jmv.27990

Faria, N. R., Mellan, T. A., Whittaker, C., Claro, I. M., Candido, D. D. S., Mishra, S., et al. (2021). Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* 372, 815–821. doi: 10.1126/science.abh2644

Focosi, D., and Casadevall, A. (2022). A critical analysis of the use of Cilgavimab plus Tixagevimab monoclonal antibody cocktail (Evusheld™) for COVID-19 prophylaxis and treatment. *Viruses* 14:1999. doi: 10.3390/v14091999

Focosi, D., and Maggi, F. (2022). Recombination in coronaviruses, with a focus on SARS-CoV-2. *Viruses* 14:1239. doi: 10.3390/v14061239

Fonager, J., Bennedbæk, M., Bager, P., Wohlfahrt, J., Ellegaard, K. M., Ingham, A. C., et al. (2022). Molecular epidemiology of the SARS-CoV-2 variant omicron BA.2 sub-lineage in Denmark, 29 November 2021 to 2 January 2022. *Euro Surveill.* 27:27. doi: 10.2807/1560-7917.es.2022.27.10.2200181

Fujino, T., Nomoto, H., Kutsuna, S., Ujiie, M., Suzuki, T., Sato, R., et al. (2021). Novel SARS-CoV-2 variant in travelers from Brazil to Japan. *Emerg. Infect. Dis.* 27, 1243–1245. doi: 10.3201/eid2704.210138

Galloway, S. E., Paul, P., MacCannell, D. R., Johansson, M. A., Brooks, J. T., MacNeil, A., et al. (2021). Emergence of SARS-CoV-2 B.1.1.7 lineage — United States, December 29, 2020–January 12, 2021. *MMWR Morb. Mortal. Wkly Rep.* 70, 95–99. doi: 10.15585/mmwr.mm7003e2

Gangavarapu, K., Latif, A. A., Mullen, J. L., Alkuzweny, M., Hufbauer, E., Tsueng, G., et al. (2023). Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nat. Methods* 20, 512–522. doi: 10.1038/s41592-023-01769-3

Garcia-Beltran, W. F., Lam, E. C., St Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., et al. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cells* 184:2523. doi: 10.1016/j.cell.2021.04.006

Ghafari, M., Liu, Q., Dhillon, A., Katzourakis, A., and Weissman, D. B. (2022). Investigating the evolutionary origins of the first three SARS-CoV-2 variants of concern. *Front. Virol.* 2:2555. doi: 10.3389/fviro.2022.942555

Gobeil, S. M.-C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., et al. (2021). D614G mutation alters SARS-CoV-2 spike conformation and enhances protease cleavage at the S1/S2 junction. *Cell Rep.* 34. doi: 10.1016/j.celrep.2020.108630

Goldman, E. (2021). How the unvaccinated threaten the vaccinated for COVID-19: a Darwinian perspective. *Proc. Natl. Acad. Sci. U. S. A.* 118:9118. doi: 10.1073/pnas.2114279118

Gomari, M. M., Tarighi, P., Choupani, E., Abkhiz, S., Mohamadzadeh, M., Rostami, N., et al. (2023). Structural evolution of Delta lineage of SARS-CoV-2. *Int. J. Biol. Macromol.* 226, 1116–1140. doi: 10.1016/j.ijbiomac.2022.11.227

Gonzalez-Reiche, A. S., Alshammary, H., Schaefer, S., Patel, G., Polanco, J., Carreño, J. M., et al. (2023). Sequential intrahost evolution and onward transmission of SARS-CoV-2 variants. *Nat. Commun.* 14:3235. doi: 10.1038/s41467-023-38867-x

Greaney, A. J., Loes, A. N., Crawford, K. H. D., Starr, T. N., Malone, K. D., Chu, H. Y., et al. (2021). Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29, 463–476.e6. doi: 10.1016/j.chom.2021.02.003

Hachmann, N. P., Miller, J., Collier, A. Y., and Barouch, D. H. (2022). Neutralization escape by SARS-CoV-2 omicron subvariant BA.4.6. *N. Engl. J. Med.* 387, 1904–1906. doi: 10.1056/NEJMc2212117

Hale, V. L., Dennis, P. M., McBride, D. S., Nolting, J. M., Madden, C., Huey, D., et al. (2022). SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature* 602, 481–486. doi: 10.1038/s41586-021-04353-x

Han, Y., An, Y., Chen, Q., Xu, K., Liu, X., Xu, S., et al. (2022). mRNA vaccines expressing homo-prototype/omicron and hetero-chimeric RBD-dimers against SARS-CoV-2. *Cell Res.* 32, 1022–1025. doi: 10.1038/s41422-022-00720-z

Harari, S., Tahor, M., Rutsinsky, N., Meijer, S., Miller, D., Henig, O., et al. (2022). Drivers of adaptive evolution during chronic SARS-CoV-2 infections. *Nat. Med.* 28, 1501–1508. doi: 10.1038/s41591-022-01882-4

Harris, E. (2023). XBB.1.16 deemed COVID-19 variant of interest. *JAMA* 329:1731. doi: 10.1001/jama.2023.7767

He, X., Hong, W., Pan, X., Lu, G., and Wei, X. (2021, 2020). SARS-CoV-2 omicron variant: characteristics and prevention. *MedComm* 2, 838–845. doi: 10.1002/mco2.110

Hill, V., du Plessis, L., Peacock, T. P., Aggarwal, D., Colquhoun, R., Carabelli, A. M., et al. (2022). The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol.* 8. doi: 10.1093/ve/veac080

Hoffmann, M., Kleine-Weber, H., and Pöhlmann, S. (2020a). A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol. Cell* 78, 779–784.e5. doi: 10.1016/j.molcel.2020.04.022

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020b). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cells* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052

Hong, Q., Han, W., Li, J., Xu, S., Wang, Y., Xu, C., et al. (2022). Molecular basis of receptor binding and antibody neutralization of omicron. *Nature* 604, 546–552. doi: 10.1038/s41586-022-04581-9

Hosch, S., Mpina, M., Nyakurungu, E., Borico, N. S., Obama, T. M. A., Ovona, M. C., et al. (2022). Genomic surveillance enables the identification of co-infections with multiple SARS-CoV-2 lineages in Equatorial Guinea. *Front. Public Health* 9:8401. doi: 10.3389/fpubh.2021.818401

Hossain, M. G., Tang, Y., Akter, S., and Zheng, C. (2021). Roles of the polybasic furin cleavage site of spike protein in SARS-CoV-2 replication, pathogenesis, and host immune responses and vaccination. *J. Med. Virol.* 94, 1815–1820. doi: 10.1002/jmv.27539

Hossen, M. L., Baral, P., Sharma, T., Gerstman, B., and Chapagain, P. (2022). Significance of the RBD mutations in the SARS-CoV-2 omicron: from spike opening to antibody escape and cell attachment. *Phys. Chem. Chem. Phys.* 24, 9123–9129. doi: 10.1039/D2CP00169A

Hou, Y. J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K. H., et al. (2020). SARS-CoV-2 D614G variant exhibits efficient replication *ex vivo* and transmission *in vivo*. *Science* 370, 1464–1468. doi: 10.1126/science.abe8499

Huang, Q., Qiu, H., Bible, P. W., Huang, Y., Zheng, F., Gu, J., et al. (2023). Early detection of SARS-CoV-2 variants through dynamic co-mutation network surveillance. *Front. Public Health* 11:5969. doi: 10.3389/fpubh.2023.1015969

Hu, B., Guo, H., Zhou, P., and Shi, Z.-L. (2020). Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* 19, 141–154. doi: 10.1038/s41579-020-00459-7

Hui, K. P. Y., Ho, J. C. W., Cheung, M., Ng, K., Ching, R. H. H., Lai, K., et al. (2022). SARS-CoV-2 omicron variant replication in human bronchus and lung *ex vivo*. *Nature* 603, 715–720. doi: 10.1038/s41586-022-04479-6

Hu, J., Peng, P., Wang, K., Fang, L., Luo, F., Jin, A., et al. (2021). Emerging SARS-CoV-2 variants reduce neutralization sensitivity to convalescent sera and monoclonal antibodies. *Cell. Mol. Immunol.* 18, 1061–1063. doi: 10.1038/s41423-021-00648-1

Imai, M., Ito, M., Kiso, M., Yamayoshi, S., Uraki, R., Fukushi, S., et al. (2023). Efficacy of antiviral agents against omicron subvariants BQ.1.1 and XBB. *N. Engl. J. Med.* 388, 89–91. doi: 10.1056/NEJMc2214302

Jackson, B., Boni, M. F., Bull, M. J., Colleran, A., Colquhoun, R. M., Darby, A. C., et al. (2021a). Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cells* 184, 5179–5188.e8. doi: 10.1016/j.cell.2021.08.014

Jackson, C. B., Farzan, M., Chen, B., and Choe, H. (2021b). Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* 23, 3–20. doi: 10.1038/s41580-021-00418-x

Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A., et al. (2021). Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* 591, 293–299. doi: 10.1038/s41586-021-03237-4

Kang, Y.-F., Sun, C., Zhuang, Z., Yuan, R.-Y., Zheng, Q., Li, J.-P., et al. (2021). Rapid development of SARS-CoV-2 spike protein receptor-binding domain self-assembled nanoparticle vaccine candidates. *ACS Nano* 15, 2738–2752. doi: 10.1021/acsnano.0c08379

Karyakarte, R. P., das, R., Dudhate, S., Agarasen, J., Pillai, P., Chandankhede, P. M., et al. (2023). Clinical characteristics and outcomes of laboratory-confirmed SARS-CoV-2 cases infected with omicron subvariants and the XBB recombinant variant. *Cureus* 15:e35261. doi: 10.7759/cureus.35261

Kelleni, M. T. (2023). Evolution of SARS CoV-2 omicron subvariants BF.7 and XBB.1.5: time to follow Africa and abort all COVID restrictions. *J. Infect.* 86:405. doi: 10.1016/j.jinf.2023.01.027

Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I. A. T. M., Gayed, S., Jahun, A., et al. (2021). SARS-CoV-2 t of chronic infection. *Nature* 592, 277–282. doi: 10.1038/s41586-021-03291-y

Kidd, M., Richter, A., Best, A., Cumley, N., Mirza, J., Percival, B., et al. (2021). S-variant SARS-CoV-2 lineage B1.1.7 is associated with significantly higher viral load in samples tested by TaqPath polymerase chain reaction. *J. Infect. Dis.* 223, 1666–1670. doi: 10.1093/infdis/jiab082

Kim, D., Quinn, J., Pinsky, B., Shah, N. H., and Brown, I. (2020). Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. *JAMA* 323:2085. doi: 10.1001/jama.2020.6266

Koopmans, M. (2021). SARS-CoV-2 and the human-animal interface: outbreaks on mink farms. *Lancet Infect. Dis.* 21, 18–19. doi: 10.1016/S1473-3099(20)30912-9

Korber, B., Fischer, W. M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., et al. (2020). Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cells* 182, 812–827.e19. doi: 10.1016/j.cell.2020.06.043

Kreier, F. (2022). Deltacron: the story of the variant that wasn't. *Nature* 602:19. doi: 10.1038/d41586-022-00149-9

Kumar, S., Thambiraja, T. S., Karuppanan, K., and Subramaniam, G. (2021). Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. *J. Med. Virol.* 94, 1641–1649. doi: 10.1002/jmv.27526

Kurhade, C., Zou, J., Xia, H., Liu, M., Chang, H. C., Ren, P., et al. (2023). Low neutralization of SARS-CoV-2 omicron BA.2.75.2, BQ.1.1 and XBB.1 by parental mRNA vaccine or a BA.5 bivalent booster. *Nat. Med.* 29, 344–347. doi: 10.1038/s41591-022-02162-x

Lacek, K. A., Rambo-Martin, B. L., Batra, D., Zheng, X., Hassell, N., Sakaguchi, H., et al. (2022a). SARS-CoV-2 Delta–omicron recombinant viruses, United States. *Emerg. Infect. Dis.* 28, 1442–1445. doi: 10.3201/eid2807.220526

Lacek, K. A., Rambo-Martin, B. L., Batra, D., Zheng, X., Sakaguchi, H., Peacock, T., et al. (2022b). Identification of a novel SARS-CoV-2 Delta-omicron recombinant virus in the United States. *bioRxiv*. 19.484981. doi: 10.1101/2022.03.19.484981

Landis, J. T., Moorad, R., Pluta, L. J., Caro-Vegas, C., McNamara, R. P., Eason, A. B., et al. (2023). Intra-host evolution provides for the continuous emergence of SARS-CoV-2 variants. *MBio* 14:e0344822. doi: 10.1128/mbio.03448-22

Lavie, M., Dubuisson, J., and Belouzard, S. (2022). SARS-CoV-2 spike Furin cleavage site and S2′ basic residues modulate the entry process in a host cell-dependent manner. *J. Virol.* 96:22. doi: 10.1128/jvi.00474-22

Le Bert, N., Clapham, H. E., Tan, A. T., Chia, W. N., Tham, C. Y. L., Lim, J. M., et al. (2021). Highly functional virus-specific cellular immune response in asymptomatic SARS-CoV-2 infection. *J. Exp. Med.* 218. doi: 10.1084/jem.20202617

Lim, S. P. (2023). Targeting SARS-CoV-2 and host cell receptor interactions. *Antivir. Res.* 210:105514. doi: 10.1016/j.antiviral.2022.105514

Lino, A., Cardoso, M. A., Martins-Lopes, P., and Gonçalves, H. M. R. (2022). Omicron – the new SARS-CoV-2 challenge? *Rev. Med. Virol.* 32:e2358. doi: 10.1002/rmv.2358

Lista, M. J., Winstone, H., Wilson, H. D., Dyer, A., Pickering, S., Galao, R. P., et al. (2022). The P681H mutation in the spike glycoprotein of the alpha variant of SARS-CoV-2 escapes IFITM restriction and is necessary for type I interferon resistance. *J. Virol.* 96:e0125022. doi: 10.1128/jvi.01250-22

Liu, Y., Liu, J., Johnson, B. A., Xia, H., Ku, Z., Schindewolf, C., et al. (2022). Delta spike P681R mutation enhances SARS-CoV-2 fitness over alpha variant. *Cell Rep.* 39:110829. doi: 10.1016/j.celrep.2022.110829

Li, X., Giorgi, E. E., Marichannegowda, M. H., Foley, B., Xiao, C., Kong, X.-P., et al. (2020a). Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci. Adv.* 6:eabb9153. doi: 10.1126/sciadv.abb9153

Li, Y.-D., Chi, W.-Y., Su, J.-H., Ferrall, L., Hung, C.-F., and Wu, T.-C. (2020b). Coronavirus vaccine development: from SARS and MERS to COVID-19. *J. Biomed. Sci.* 27:695. doi: 10.1186/s12929-020-00695-2

Looi, M.-K. (2023). What do we know about the Arcturus XBB.1.16 subvariant? *Stroke Vasc. Neurol.*:p1074. doi: 10.1136/bmj.p1074

Lubinski, B., Fernandes, M. H. V., Frazier, L., Tang, T., Daniel, S., Diel, D. G., et al. (2022). Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (alpha) spike. *iScience* 25:103589. doi: 10.1016/j.isci.2021.103589

Lukassen, S., Chua, R. L., Trefzer, T., Kahn, N. C., Schneider, M. A., Muley, T., et al. (2020). SARS-CoV-2 receptor ACE 2 and TMPRSS 2 are primarily expressed in bronchial transient secretory cells. *EMBO J.* 39:e105114. doi: 10.15252/embj.20105114

Lu, L., Sikkema, R., Velkers, F., Nieuwenhuijse, D., Fischer, E., Meijer, P., et al. (2021). Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat. Commun.* 12:6802. doi: 10.1038/s41467-021-27096-9

Lu, X., Chen, Y., and Zhang, G. (2023). Functional evolution of SARS-CoV-2 spike protein: maintaining wide host spectrum and enhancing infectivity via surface charge of spike protein. *Comput. Struct. Biotechnol. J.* 21, 2068–2074. doi: 10.1016/j.csbj.2023.03.010

Magazine, N., Zhang, T., Wu, Y., McGee, M. C., Veggiani, G., and Huang, W. (2022). Mutations and evolution of the SARS-CoV-2 spike protein. *Viruses* 14:640. doi: 10.3390/v14030640

Magiorkinis, G. (2023). On the evolution of SARS-CoV-2 and the emergence of variants of concern. *Trends Microbiol.* 31, 5–8. doi: 10.1016/j.tim.2022.10.008

Mahase, E. (2021). Delta variant: what is happening with transmission, hospital admissions, and restrictions? *Stroke Vasc. Neurol.* 373:n1513. doi: 10.1136/bmj.n1513

Mahase, E. (2022). COVID-19: what do we know about the delta omicron recombinant variant? *Stroke Vasc. Neurol.* 376:o792. doi: 10.1136/bmj.o792

Mahase, E. (2023). Covid-19: what do we know about XBB.1.5 and should we be worried? *Stroke Vasc. Neurol.* 380:p153. doi: 10.1136/bmj.p153

Ma, K. C., Shirk, P., Lambrou, A. S., Hassell, N., Zheng, X., Payne, A. B., et al. (2023). Genomic surveillance for SARS-CoV-2 variants: circulation of omicron lineages — United States, January 2022–May 2023. *MMWR Morb. Mortal. Wkly Rep.* 72, 651–656. doi: 10.15585/mmwr.mm7224a2

Mallapaty, S. (2022). Where did omicron come from? Three key theories. *Nature* 602, 26–28. doi: 10.1038/d41586-022-00215-2

Ma, M.-J., Qiu, S.-F., Cui, X.-M., Ni, M., Liu, H.-J., Ye, R.-Z., et al. (2022a). Persistent SARS-CoV-2 infection in asymptomatic young adults. *Sig. Transduct. Target Ther.* 7:77. doi: 10.1038/s41392-022-00931-1

Mansbach, R. A., Chakraborty, S., Nguyen, K., Montefiori, D. C., Korber, B., and Gnanakaran, S. (2021). The SARS-CoV-2 spike variant D614G favors an open conformational state. *Sci. Adv.* 7:3671. doi: 10.1126/sciadv.abf3671

Marques, A. D., Sherrill-Mix, S., Everett, J. K., Adhikari, H., Reddy, S., Ellis, J. C., et al. (2022). Multiple introductions of SARS-CoV-2 alpha and Delta variants into white-tailed deer in Pennsylvania. *MBio* 13:e0210122. doi: 10.1128/mbio.02101-22

Martin, D. P., Lytras, S., Lucaci, A. G., Maier, W., Grüning, B., Shank, S. D., et al. (2022). Selection analysis identifies clusters of unusual mutational changes in omicron lineage BA.1 that likely impact spike function. *Mol. Biol. Evol.* 39. doi: 10.1093/molbev/msac061

Maulud, S. Q., Hasan, D. A., Ali, R. K., Rashid, R. F., Saied, A. A., Dhawan, M., et al. (2022). Deltacron: apprehending a new phase of the COVID-19 pandemic. *Int. J. Surg.* 102:106654. doi: 10.1016/j.ijsu.2022.106654

Ma, W., Yang, J., Fu, H., Su, C., Yu, C., Wang, Q., et al. (2022b). Genomic perspectives on the emerging SARS-CoV-2 omicron variant. *Genom. Proteom. Bioinformat.* 20, 60–69. doi: 10.1016/j.gpb.2022.01.001

Meng, B., Abdullahi, A., Ferreira, I. A. T. M., Goonawardane, N., Saito, A., Kimura, I., et al. (2022). Altered TMPRSS2 usage by SARS-CoV-2 omicron impacts infectivity and fusogenicity. *Nature* 603, 706–714. doi: 10.1038/s41586-022-04474-x

Milne, G., Hames, T., Scotton, C., Gent, N., Johnsen, A., Anderson, R. M., et al. (2021). Does infection with or vaccination against SARS-CoV-2 lead to lasting immunity? *Medicine* 9, 1450–1466. doi: 10.1016/S2213-2600(21)00407-0

Mlcochova, P., Kemp, S. A., Dhar, M. S., Papa, G., Meng, B., Ferreira, I. A. T. M., et al. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* 599, 114–119. doi: 10.1038/s41586-021-03944-y

Moelling, K. (2021). Within-host and between-host evolution in SARS-CoV-2—new Variant's source. *Viruses* 13:751. doi: 10.3390/v13050751

Moisan, A., Mastrovito, B., de Oliveira, F., Martel, M., Hedin, H., Leoz, M., et al. (2022). Evidence of transmission and circulation of Deltacron XD recombinant severe acute respiratory syndrome coronavirus 2 in Northwest France. *Clin. Infect. Dis.* 75, 1841–1844. doi: 10.1093/cid/ciac360

Motozono, C., Toyoda, M., Zahradnik, J., Saito, A., Nasser, H., Tan, T. S., et al. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* 29, 1124–1136.e11. doi: 10.1016/j.chom.2021.06.006

Müller, N. F., Kistler, K. E., and Bedford, T. (2022). A Bayesian approach to infer recombination patterns in coronaviruses. *Nat. Commun.* 13:31749. doi: 10.1038/s41467-022-31749-8

Nehlmeier, I., Kempf, A., Arora, P., Cossmann, A., Dopfer-Jablonka, A., Stankov, M. V., et al. (2023). Host cell entry and neutralisation sensitivity of the SARS-CoV-2 XBB.1.16 lineage. *Cell. Mol. Immunol.* 8, 1–3. doi: 10.1038/s41423-023-01030-z

Nguyen, H. T., Zhang, S., Wang, Q., Anang, S., Wang, J., Ding, H., et al. (2021). Spike glycoprotein and host cell determinants of SARS-CoV-2 entry and cytopathic effects. *J. Virol.* 95. doi: 10.1128/JVI.02304-20

O'Toole, Á., Pybus, O. G., Abram, M. E., Kelly, E. J., and Rambaut, A. (2022). Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics* 23:8358. doi: 10.1186/s12864-022-08358-2

Ou, J., Lan, W., Wu, X., Zhao, T., Duan, B., Yang, P., et al. (2022). Tracking SARS-CoV-2 omicron diverse spike gene mutations identifies multiple inter-variant recombination events. *Sig. Transduct. Target Ther.* 7:138. doi: 10.1038/s41392-022-00992-2

Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., et al. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* 12:848. doi: 10.1038/s41467-021-21118-2

Pang, Y. T., Acharya, A., Lynch, D. L., Pavlova, A., and Gumbart, J. C. (2022). SARS-CoV-2 spike opening dynamics and energetics reveal the individual roles of glycans and their collective impact. *Commun. Biol.* 5:4138. doi: 10.1038/s42003-022-04138-6

Pan, Y., Wang, L., Feng, Z., Xu, H., Li, F., Shen, Y., et al. (2023). Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis. *Lancet* 401, 664–672. doi: 10.1016/S0140-6736(23)00129-0

Patiño-Galindo, J. Á., Filip, I., and Rabadan, R. (2021). Global patterns of recombination across human viruses. *Mol. Biol. Evol.* 38, 2520–2531. doi: 10.1093/molbev/msab046

Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L., Frise, R., Swann, O. C., et al. (2021). The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol.* 6, 899–909. doi: 10.1038/s41564-021-00908-w

Pedro, N., Silva, C. N., Magalhães, A. C., Cavadas, B., Rocha, A. M., Moreira, A. C., et al. (2021). Dynamics of a dual SARS-CoV-2 lineage co-infection on a prolonged viral shedding COVID-19 case: insights into clinical severity and disease duration. *Microorganisms* 9:300. doi: 10.3390/microorganisms9020300

Perales, C. (2021). Mutation rates, mutation frequencies, and proofreading-repair activities in RNA virus genetics. *Viruses* 13:1882. doi: 10.3390/v13091882

Perez-Florido, J., Casimiro-Soriguer, C. S., Ortuño, F., Fernandez-Rueda, J. L., Aguado, A., Lara, M., et al. (2023). Detection of high level of co-infection and the emergence of novel SARS CoV-2 Delta-omicron and omicron-omicron recombinants in the epidemiological surveillance of Andalusia. *IJMS* 24:2419. doi: 10.3390/ijms24032419

Planas, D., Bruel, T., Staropoli, I., Guivel-Benhassine, F., Porrot, F., Maes, P., et al. (2023). Resistance of omicron subvariants BA.2.75.2, BA.4.6, and BQ.1.1 to neutralizing antibodies. *Nat. Commun.* 14:824. doi: 10.1038/s41467-023-36561-6

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 592, 116–121. doi: 10.1038/s41586-020-2895-3

Qu, P., Evans, J. P., Faraone, J. N., Zheng, Y.-M., Carlin, C., Anghelina, M., et al. (2023). Enhanced neutralization resistance of SARS-CoV-2 omicron subvariants BQ.1, BQ.1.1, BA.4.6, BF.7, and BA.2.75.2. *Cell Host Microbe* 31, 9–17.e3. doi: 10.1016/j.chom.2022.11.012

Qu, P., Evans, J. P., Zheng, Y.-M., Carlin, C., Saif, L. J., Oltz, E. M., et al. (2022). Evasion of neutralizing antibody responses by the SARS-CoV-2 BA.2.75 variant. *Cell Host Microbe* 30, 1518–1526.e4. doi: 10.1016/j.chom.2022.09.015

Rahbar Saadat, Y., Hosseiniyan Khatibi, S. M., Zununi Vahed, S., and Ardalan, M. (2021). Host serine proteases: a potential targeted therapy for COVID-19 and influenza. *Front. Mol. Biosci.* 8:5528. doi: 10.3389/fmolb.2021.725528

Rahimi, F., and Bezmin Abadi, A. T. (2022a). The omicron subvariant BA.2: birth of a new challenge during the COVID-19 pandemic. *Int. J. Surg.* 99:106261. doi: 10.1016/j.ijsu.2022.106261

Rahimi, F., and Bezmin Abadi, A. T. (2022b). Detection of the XE subvariant of SARS-CoV-2: a perspective. *Int. J. Surg.* 101:106642. doi: 10.1016/j.ijsu.2022.106642

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5

Ren, W., Zhang, Y., Rao, J., Wang, Z., Lan, J., Liu, K., et al. (2023). Evolution of immune evasion and host range expansion by the SARS-CoV-2 B.1.1.529 (omicron) variant. *M bio* 14. doi: 10.1128/mbio.00416-23

Robson, F., Khan, K. S., le, T. K., Paris, C., Demirbag, S., Barfuss, P., et al. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell* 79, 710–727. doi: 10.1016/j.molcel.2020.07.027

Rockett, R. J., Draper, J., Gall, M., Sim, E. M., Arnott, A., Agius, J. E., et al. (2022). Co-infection with SARS-CoV-2 omicron and Delta variants revealed by genomic surveillance. *Nat. Commun.* 13:2745. doi: 10.1038/s41467-022-30518-x

Sabir, J. S. M., Lam, T. T.-Y., Ahmed, M. M. M., Li, L., Shen, Y., Abo-Aba, S. E. M., et al. (2016). Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* 351, 81–84. doi: 10.1126/science.aac8608

Saito, A., Irie, T., Suzuki, R., Maemura, T., Nasser, H., Uriu, K., et al. (2021). Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* 602, 300–306. doi: 10.1038/s41586-021-04266-9

Saito, A., Tamura, T., Zahradnik, J., Deguchi, S., Tabata, K., Anraku, Y., et al. (2022). Virological characteristics of the SARS-CoV-2 omicron BA.2.75 variant. *Cell Host Microbe* 30, 1540–1555.e15. doi: 10.1016/j.chom.2022.10.003

Saldivar-Espinoza, B., Garcia-Segura, P., Novau-Ferré, N., Macip, G., Martínez, R., Puigbò, P., et al. (2023). The mutational landscape of SARS-CoV-2. *IJMS* 24:9072. doi: 10.3390/ijms24109072

Samoilov, A. E., Kaptelova, V. V., Bukharina, A. Y., Shipulina, O. Y., Korneenko, E. V., Saenko, S. S., et al. (2021). Case report: change of dominant strain during dual SARS-CoV-2 infection. *BMC Infect. Dis.* 21:959. doi: 10.1186/s12879-021-06664-w

Sathyaseelan, C., Magateshvaren Saras, M. A., Prasad Patro, L. P., Uttamrao, P. P., and Rathinavelan, T. (2023). CoVe-tracker: an interactive SARS-CoV-2 Pan proteome evolution tracker. *J. Proteome Res.* 22, 1984–1996. doi: 10.1021/acs.jproteome.3c00068

Saxena, S. K., Kumar, S., Ansari, S., Paweska, J. T., Maurya, V. K., Tripathi, A. K., et al. (2022). Characterization of the novel SARS-CoV-2 omicron (B.1.1.529) variant of concern and its global perspective. *J. Med. Virol.* 94, 1738–1744. doi: 10.1002/jmv.27524

Scarpa, F., Giovanetti, M., Azzena, I., Locci, C., Casu, M., Fiori, P. L., et al. (2023a). Genome-based survey of the SARS-CoV-2 BF.7 variant from Asia. *J. Med. Virol.* 95:e28714. doi: 10.1002/jmv.28714

Scarpa, F., Sanna, D., Azzena, I., Casu, M., Cossu, P., Fiori, P. L., et al. (2023b). Genome-based comparison between the recombinant SARS-CoV-2 XBB and its parental lineages. *J. Med. Virol.* 95:e28625. doi: 10.1002/jmv.28625

Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., et al. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11727–11734. doi: 10.1073/pnas.2003138117

Shapira, G., Patalon, T., Gazit, S., and Shomron, N. (2023). Immunosuppression as a hub for SARS-CoV-2 mutational drift. *Viruses* 15:855. doi: 10.3390/v15040855

Shen, J., Fan, J., Zhao, Y., Jiang, D., Niu, Z., Zhang, Z., et al. (2023). Innate and adaptive immunity to SARS-CoV-2 and predisposing factors. *Front. Immunol.* 14:9326. doi: 10.3389/fimmu.2023.1159326

Sheward, D. J., Kim, C., Fischbach, J., Sato, K., Muschiol, S., Ehling, R. A., et al. (2022). Omicron sublineage BA.2.75.2 exhibits extensive escape from neutralising antibodies. *Lancet Infect. Dis.* 22, 1538–1540. doi: 10.1016/S1473-3099(22)00663-6

Shi, R., Shan, C., Duan, X., Chen, Z., Liu, P., Song, J., et al. (2020). A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* 584, 120–124. doi: 10.1038/s41586-020-2381-y

Singh, J., Vashishtha, S., Rahman, S. A., Ehtesham, N. Z., Alam, A., Kundu, B., et al. (2022). Energetics of spike protein opening of SARS-CoV-1 and SARS-CoV-2 and its variants of concern: implications in host receptor scanning and transmission. *Biochemistry* 61, 2188–2197. doi: 10.1021/acs.biochem.2c00301

Smith, C. A., and Ashby, B. (2022). Antigenic evolution of SARS-CoV-2 in immunocompromised hosts. *EMPH* 11, 90–100. doi: 10.1093/emph/eoac037

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H. D., Dingens, A. S., et al. (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cells* 182, 1295–1310.e20. doi: 10.1016/j.cell.2020.08.012

Sungnak, W., Huang, N., Bécavin, C., Berg, M., Queen, R., Litvinukova, M., et al. (2020). SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* 26, 681–687. doi: 10.1038/s41591-020-0868-6

Sztain, T., Ahn, S.-H., Bogetti, A. T., Casalino, L., Goldsmith, J. A., Seitz, E., et al. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* 13, 963–968. doi: 10.1038/s41557-021-00758-3

Tamura, T., Ito, J., Uriu, K., Zahradnik, J., Kida, I., Nasser, H., et al. (2022). Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two omicron subvariants. *bioRxiv*. 521986. doi: 10.1101/2022.12.27.521986

Tan, C. Y., Chiew, C. J., Pang, D., Lee, V. J., Ong, B., Lye, D. C., et al. (2023). Protective immunity of SARS-CoV-2 infection and vaccines against medically attended symptomatic omicron BA.4, BA.5, and XBB reinfections in Singapore: a national cohort study. *Lancet Infect. Dis.* 23, 799–805. doi: 10.1016/S1473-3099(23)00060-9

Tao, K., Tzou, P. L., Nouhin, J., Gupta, R. K., de Oliveira, T., Kosakovsky Pond, S. L., et al. (2021). The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* 22, 757–773. doi: 10.1038/s41576-021-00408-x

Tegally, H., Moir, M., Everatt, J., Giovanetti, M., Scheepers, C., Wilkinson, E., et al. (2022). Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in South Africa. *Nat. Med.* 28, 1785–1790. doi: 10.1038/s41591-022-01911-2

Tegally, H., Wilkinson, E., Lessells, R. J., Giandhari, J., Pillay, S., Msomi, N., et al. (2021). Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* 27, 440–446. doi: 10.1038/s41591-021-01255-3

Thorne, L. G., Bouhaddou, M., Reuschl, A.-K., Zuliani-Alvarez, L., Polacco, B., Pelin, A., et al. (2021). Evolution of enhanced innate immune evasion by SARS-CoV-2. *Nature* 602, 487–495. doi: 10.1038/s41586-021-04352-y

Tian, D., Sun, Y., Xu, H., and Ye, Q. (2022). The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 omicron variant. *J. Med. Virol.* 94, 2376–2383. doi: 10.1002/jmv.27643

Tian, J.-H., Patel, N., Haupt, R., Zhou, H., Weston, S., Hammond, H., et al. (2021). SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 immunogenicity in baboons and protection in mice. *Nat. Commun.* 12:372. doi: 10.1038/s41467-020-20653-8

Tortorici, M. A., Beltramello, M., Lempp, F. A., Pinto, D., Dang, H. V., Rosen, L. E., et al. (2020). Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* 370, 950–957. doi: 10.1126/science.abe3354

Tsueng, G., Mullen, J. L., Alkuzweny, M., Cano, M., Rush, B., Haag, E., et al. (2023). Outbreak.info research library: a standardized, searchable platform to discover and explore COVID-19 resources. *Nat. Methods* 20, 536–540. doi: 10.1038/s41592-023-01770-w

Turakhia, Y., Thornlow, B., Hinrichs, A., McBroome, J., Ayala, N., Ye, C., et al. (2022). Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. *Nature* 609, 994–997. doi: 10.1038/s41586-022-05189-9

Turelli, P., Fenwick, C., Raclot, C., Genet, V., Pantaleo, G., and Trono, D. (2022). P2G3 human monoclonal antibody neutralizes SARS-CoV-2 omicron subvariants including BA.4 and BA.5 and Bebtelovimab escape mutants. *bioRxiv*. 501852. doi: 10.1101/2022.07.28.501852

Uraki, R., Ito, M., Kiso, M., Yamayoshi, S., Iwatsuki-Horimoto, K., Sakai-Tagawa, Y., et al. (2023). Efficacy of antivirals and bivalent mRNA vaccines against SARS-CoV-2 isolate CH.1.1. *Lancet Infect. Dis.* 23, 525–526. doi: 10.1016/S1473-3099(23)00132-9

van Egeren, D., Stoddard, M., White, L. F., Hochberg, N. S., Rogers, M. S., Zetter, B., et al. (2023). Vaccines alone cannot slow the evolution of SARS-CoV-2. *Vaccines* 11:853. doi: 10.3390/vaccines11040853

Varabyou, A., Pockrandt, C., Salzberg, S. L., and Pertea, M. (2021). Rapid detection of inter-clade recombination in SARS-CoV-2 with Bolotie. *Genetics* 218:74. doi: 10.1093/genetics/iyab074

Varghese, R., Pai, S., Kumar, D., and Sharma, R. (2023). SARS-CoV-2 XBB.1.16 variant: India in focus? *J. Med. Virol.* 95:28829. doi: 10.1002/jmv.28829

Venkatakrishnan, A. J., Anand, P., Lenehan, P. J., Suratekar, R., Raghunathan, B., Niesen, M. J. M., et al. (2022). On the origins of Omicron's unique spike gene insertion. *Vaccines* 10:1509. doi: 10.3390/vaccines10091509

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., et al. (2022). Rapid epidemic expansion of the SARS-CoV-2 omicron variant in southern Africa. *Nature* 603, 679–686. doi: 10.1038/s41586-022-04411-y

Vkovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2020). Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19, 155–170. doi: 10.1038/s41579-020-00468-6

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., et al. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 593, 266–269. doi: 10.1038/s41586-021-03470-x

Walls, A. C., Park, Y.-J., Tortorici, M. A., Wall, A., McGuire, A. T., and Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cells* 181, 281–292.e6. doi: 10.1016/j.cell.2020.02.058

Wang, C., Liu, B., Zhang, S., Huang, N., Zhao, T., Lu, Q., et al. (2022). Differences in incidence and fatality of COVID-19 by SARS-CoV-2 omicron variant versus Delta variant in relation to vaccine coverage: a world-wide review. *J. Med. Virol.* 95:e28118. doi: 10.1002/jmv.28118

Wang, L., and Gao, G. F. (2022). The "wolf" is indeed coming: recombinant "Deltacron" SARS-CoV-2 detected. *China. CDC. Wkly.* 4, 285–287. doi: 10.46234/ccdcw2022.054

Wang, L., Zhou, H., Li, J., Cheng, Y., Zhang, S., Aliyari, S., et al. (2022). Potential intervariant and intravariant recombination of Delta and omicron variants. *J. Med. Virol.* 94, 4830–4838. doi: 10.1002/jmv.27939

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* 593, 130–135. doi: 10.1038/s41586-021-03398-2

Wang, Q., Iketani, S., Li, Z., Guo, Y., Yeh, A. Y., Liu, M., et al. (2022a). Antigenic characterization of the SARS-CoV-2 omicron subvariant BA.2.75. *Cell Host Microbe* 30, 1512–1517.e4. doi: 10.1016/j.chom.2022.09.002

Wang, Q., Iketani, S., Li, Z., Liu, L., Guo, Y., Huang, Y., et al. (2022b). Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. *Cells* 186, 279–286.e8. doi: 10.1016/j.cell.2022.12.018

Wang, Q., Li, Z., Ho, J., Guo, Y., Yeh, A. Y., Mohri, H., et al. (2022c). Resistance of SARS-CoV-2 omicron subvariant BA.4.6 to antibody neutralisation. *Lancet Infect. Dis.* 22, 1666–1668. doi: 10.1016/s1473-3099(22)00694-6

Wang, S., Xu, X., Wei, C., Li, S., Zhao, J., Zheng, Y., et al. (2021). Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. *J. Med. Virol.* 94, 310–317. doi: 10.1002/jmv.27331

Wang, Y., Liu, D., Shi, W., Lu, R., Wang, W., Zhao, Y., et al. (2015). Origin and possible genetic recombination of the Middle East respiratory syndrome coronavirus from the first imported case in China: Phylogenetics and coalescence analysis. *M bio* 6, e01280–e01215. doi: 10.1128/mBio.01280-15

Wassenaar, T. M., Wanchai, V., Buzard, G., and Ussery, D. W. (2022). The first three waves of the Covid-19 pandemic hint at a limited genetic repertoire for SARS-CoV-2. *FEMS Microbiol. Rev.* 46:fuac003. doi: 10.1093/femsre/fuac003

Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S., and Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*:eabb9983. doi: 10.1126/science.abb9983

Watanabe, Y., Bowden, T. A., Wilson, I. A., and Crispin, M. (2019). Exploitation of glycosylation in enveloped virus pathobiology. *Biochim. Biophys. Acta Gen. Subj.* 1863, 1480–1497. doi: 10.1016/j.bbagen.2019.05.012

Weissman, D., Alameh, M.-G., de Silva, T., Collini, P., Hornsby, H., Brown, R., et al. (2020). D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. *Cell Host Microbe* 29, 23–31.e4. doi: 10.1016/j.chom.2020.11.012

Wertheim, J. O., Wang, J. C., Leelawong, M., Martin, D. P., Havens, J. L., Chowdhury, M. A., et al. (2022). Capturing intrahost recombination of SARS-CoV-2 during superinfection with alpha and epsilon variants in New York City. *medRxiv*. 22269300. doi: 10.1101/2022.01.18.22269300

Wibmer, C. K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Oosthuysen, B., et al. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by south African COVID-19 donor plasma. *Nat. Med.* 27, 622–625. doi: 10.1038/s41591-021-01285-x

Wilkinson, S. A. J., Richter, A., Casey, A., Osman, H., Mirza, J. D., Stockton, J., et al. (2022). Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evol.* 8. doi: 10.1093/ve/veac050

Willett, B. J., Grove, J., MacLean, O. A., Wilkie, C., De Lorenzo, G., Furnon, W., et al. (2022). SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat. Microbiol.* 7, 1161–1179. doi: 10.1038/s41564-022-01143-7

Wolf, J. M., Wolf, L. M., Bello, G. L., Maccari, J. G., and Nasi, L. A. (2022). Molecular evolution of SARS-CoV-2 from December 2019 to august 2022. *J. Med. Virol.* 95:28366. doi: 10.1002/jmv.28366

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., et al. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263. doi: 10.1126/science.abb2507

Wu, L., Zhou, L., Mo, M., Liu, T., Wu, C., Gong, C., et al. (2022). SARS-CoV-2 omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Sig. Transduct. Target Ther.* 7:8. doi: 10.1038/s41392-021-00863-2

Xia, S., Jiao, F., Wang, L., Yu, X., Lu, T., Fu, Y., et al. (2023a). SARS-CoV-2 omicron XBB subvariants exhibit enhanced fusogenicity and substantial immune evasion in elderly population, but high sensitivity to pan-coronavirus fusion inhibitors. *J. Med. Virol.* 95:e28641. doi: 10.1002/jmv.28641

Xia, S., Liu, Z., and Jiang, S. (2023b). High-throughput screening of mutations affecting SARS-CoV-2 spike functions. *Trends Immunol.* 44, 321–323. doi: 10.1016/j.it.2023.03.010

Xia, S., Wang, L., Zhu, Y., Lu, L., and Jiang, S. (2022). Origin, virological features, immune evasion and intervention of SARS-CoV-2 omicron sublineages. *Sig. Transduct. Target Ther.* 7:1105. doi: 10.1038/s41392-022-01105-9

Yamasoba, D., Uriu, K., Plianchaisuk, A., Kosugi, Y., Pan, L., Zahradnik, J., et al. (2023). Virological characteristics of the SARS-CoV-2 Omicron XBB.1.16 variant. *Lancet* 23, 527–528. doi: 10.1101/2023.04.06.535883

Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., et al. (2020a). A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature* 586, 572–577. doi: 10.1038/s41586-020-2599-8

Yang, P.-H., Ding, Y.-B., Xu, Z., Pu, R., Li, P., Yan, J., et al. (2020b). Increased circulating level of interleukin-6 and CD8+ T cell exhaustion are associated with progression of COVID-19. *Infect. Dis. Poverty* 9:161. doi: 10.1186/s40249-020-00780-6

Yang, S., Li, Y., Dai, L., Wang, J., He, P., Li, C., et al. (2021). Safety and immunogenicity of a recombinant tandem-repeat dimeric RBD-based protein subunit vaccine (ZF2001) against COVID-19 in adults: two randomised, double-blind, placebo-controlled, phase 1 and 2 trials. *Lancet Infect. Dis.* 21, 1107–1119. doi: 10.1016/S1473-3099(21)00127-4

Ye, G., Liu, B., and Li, F. (2022). Cryo-EM structure of a SARS-CoV-2 omicron spike protein ectodomain. *Nat. Commun.* 13:28882. doi: 10.1038/s41467-022-28882-9

Yewdell, J. W. (2021). Antigenic drift: Understanding COVID-19. *Immunity* 54, 2681–2687. doi: 10.1016/j.immuni.2021.11.016

Yin, W., Xu, Y., Xu, P., Cao, X., Wu, C., Gu, C., et al. (2022). Structures of the omicron spike trimer with ACE2 and an anti-omicron antibody. *Science* 375, 1048–1053. doi: 10.1126/science.abn8863

Yue, C., Song, W., Wang, L., Jian, F., Chen, X., Gao, F., et al. (2023). ACE2 binding and antibody evasion in enhanced transmissibility of XBB.1.5. *Lancet Infect. Dis.* 23, 278–280. doi: 10.1016/S1473-3099(23)00010-5

Yurkovetskiy, L., Wang, X., Pascal, K. E., Tomkins-Tinch, C., Nyalile, T. P., Wang, Y., et al. (2020). Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cells* 183, 739–751.e8. doi: 10.1016/j.cell.2020.09.032

Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., et al. (2021). SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD *in vitro* evolution. *Nat. Microbiol.* 6, 1188–1198. doi: 10.1038/s41564-021-00954-4

Zhang, J., Cai, Y., Xiao, T., Lu, J., Peng, H., Sterling, S. M., et al. (2021a). Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 372, 525–530. doi: 10.1126/science.abf2303

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., et al. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nat. Commun.* 11:6013. doi: 10.1038/s41467-020-19808-4

Zhang, L., Li, Q., Liang, Z., Li, T., Liu, S., Cui, Q., et al. (2021b). The significant immune escape of pseudotyped SARS-CoV-2 variant omicron. *Emerg. Microbes. Infect.* 11, 1–5. doi: 10.1080/22221751.2021.2017757

Zhang, L., Mann, M., Syed, Z. A., Reynolds, H. M., Tian, E., Samara, N. L., et al. (2021c). Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* 118. doi: 10.1073/pnas.2109905118

Zhang, Y., Ndzouboukou, J. B., Lin, X., Hou, H., Wang, F., Yuan, L., et al. (2022). SARS-CoV-2 evolves to reduce but not abolish neutralizing action. *J. Med. Virol.* 95:28207. doi: 10.1002/jmv.28207

Zhao, H., Lu, L., Peng, Z., Chen, L.-L., Meng, X., Zhang, C., et al. (2022). SARS-CoV-2 omicron variant shows less efficient replication and fusion activity when compared with Delta variant in TMPRSS2-expressed cells. *Emerg. Microbes. Infect.* 11, 277–283. doi: 10.1080/22221751.2021.2023329

Zhou, B., Thao, T. T. N., Hoffmann, D., Taddeo, A., Ebert, N., Labroussaa, F., et al. (2021). SARS-CoV-2 spike D614G change enhances replication and transmission. *Nature* 592, 122–127. doi: 10.1038/s41586-021-03361-1

Zhou, Y., Zhi, H., and Teng, Y. (2022). The outbreak of SARS-CoV-2 omicron lineages, immune escape, and vaccine effectivity. *J. Med. Virol.* 95:28138. doi: 10.1002/jmv.28138

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 08 August 2023
DOI 10.3389/fmicb.2023.1222301

OPEN ACCESS

EDITED BY
Axel Cloeckaert,
Institut National de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE), France

REVIEWED BY
Richard Johnathan Orton,
University of Glasgow, United Kingdom
Jayanthi S. Shastri,
Topiwala National Medical College and BYL Nair Charitable Hospital, India

*CORRESPONDENCE
Hezhao Ji
✉ hezhao.ji@phac-aspc.gc.ca

RECEIVED 17 May 2023
ACCEPTED 17 July 2023
PUBLISHED 08 August 2023

CITATION
Li K, Melnychuk S, Sandstrom P and Ji H (2023) Tracking the evolution of the SARS-CoV-2 Delta variant of concern: analysis of genetic diversity and selection across the whole viral genome. *Front. Microbiol.* 14:1222301. doi: 10.3389/fmicb.2023.1222301

Tracking the evolution of the SARS-CoV-2 Delta variant of concern: analysis of genetic diversity and selection across the whole viral genome

Katherine Li[1,2], Stephanie Melnychuk[1], Paul Sandstrom[1,2] and Hezhao Ji[1,2]*

[1]National Microbiology Laboratory at JC Wilt Infectious Diseases Research Centre, Public Health Agency of Canada, Winnipeg, MB, Canada, [2]Department of Medical Microbiology and Infectious Diseases, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB, Canada

Background: Since 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has diversified extensively, producing five highly virulent lineages designated as variants of concern (VOCs). The Delta VOC emerged in India with increased transmission, immune evasion, and mortality, causing a massive global case surge in 2021. This study aims to understand how the Delta VOC evolved by characterizing mutation patterns in the viral population before and after its emergence. Furthermore, we aim to identify the influence of positive and negative selection on VOC evolution and understand the prevalence of different mutation types in the viral genome.

Methods: Three groups of whole viral genomes were retrieved from GISAID, sourced from India, with collection periods as follows: Group A—during the initial appearance of SARS-CoV-2; Group B—just before the emergence of the Delta variant; Group C—after the establishment of the Delta variant in India. Mutations in >1% of each group were identified with BioEdit to reveal differences in mutation quantity and type. Sites under positive or negative selection were identified with FUBAR. The results were compared to determine how mutations correspond with selective pressures and how viral mutation profiles changed to reflect genetic diversity before and after VOC emergence.

Results: The number of mutations increased progressively in Groups A–C, with Group C reporting a 2.2- and 1.9-fold increase from Groups A and B, respectively. Among all the observed mutations, Group C had the highest percentage of deletions (22.7%; vs. 4.2% and 2.6% in Groups A and B, respectively), and most mutations altered the final amino acid code, such as non-synonymous substitutions and deletions. Conversely, Group B had the most synonymous substitutions that are effectively silent. The number of sites experiencing positive selection increased in Groups A–C, but Group B had 2.4- and 2.6 times more sites under negative selection compared to Groups A and C, respectively.

Conclusion: Our findings demonstrated that viral genetic diversity continuously increased during and after the emergence of the Delta VOC. Despite this, Group B reports heightened negative selection, which potentially preserves important gene regions during evolution. Group C contains an unprecedented quantity of mutations and positively selected sites, providing strong evidence of active viral adaptation in the population.

KEYWORDS

SARS-CoV-2, evolution, genetic diversity, selection, whole genome

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a betacoronavirus infamous for causing the coronavirus disease 2019 (COVID-19), emerged in late 2019 and rapidly escalated into a global pandemic. The SARS-CoV-2 genome spans approximately 30 kb in length and encodes 16 non-structural, four structural, and nine accessory proteins (Wu et al., 2020; Bai et al., 2022). The structural proteins include the nucleocapsid (N), membrane (M), envelope (E), and spike (S) (Khailany et al., 2020). Glycosylated S proteins cover the virion surface and contain the receptor-binding domain (RBD), which mediates host-binding interactions (Huang et al., 2020; Letko et al., 2020). The E protein contributes to the formation of the viral envelope, while the M protein plays an essential role in virion assembly (Bai et al., 2022). The N protein binds the viral genome and has multiple roles in RNA synthesis and translation, viral replication, and cell cycle regulation (Bai et al., 2022). The ORF1ab gene encodes 16 non-structural proteins that are essential for viral RNA replication and transcription (Bai et al., 2022). The accessory proteins encoded by other open reading frames (ORFs) encompass a diverse range of functions, including viral release, inhibition of host cellular functions and immune response, formation of ion channels, and interactions with other viral proteins (Li et al., 2020; Toft-Bertelsen et al., 2021; Zinzula, 2021; Bai et al., 2022).

SARS-CoV-2, such as other RNA viruses, exhibits a high mutation rate primarily attributed to its error-prone RNA-dependent RNA polymerase (Duarte et al., 2022). When coupled with high rates of viral reproduction, these mutations may occur throughout the viral genome, allowing viral quasi-species to arise and persist within the infected host (Karamitros et al., 2020). Additionally, exposure to an enormous pool of susceptible hosts favors the rapid evolution of SARS-CoV-2, producing numerous viral variants that circulate within human populations (Duarte et al., 2022). This has been well illustrated by the many SARS-CoV-2 variants of concern (VOCs) designated by the World Health Organization (WHO), which possess mutations that confer high-risk traits including increased transmissibility, increased virulence, and reduced susceptibility to vaccines and therapeutics (World Health Organization, 2021). The designated VOCs included the Alpha, Beta, Gamma, Delta, and Omicron lineages (World Health Organization, 2021). In particular, the Delta VOC has been known for its increased virulence, pathogenicity, and severity of COVID-19 disease (Liu and Rocklöv, 2021; Zhang et al., 2022).

The first SARS-CoV-2 infection case in India was reported in Kerala on 30 January 2020 (Andrews et al., 2020). On 11 March 2020, the WHO declared COVID-19 a global pandemic, prompting the Indian government to impose a nationwide lockdown on 25 March 2020 (Siddiqui et al., 2020). A subsequent wave of transmission hit India during the Spring of 2021, driven by the emergence of a new variant, B.1.617.2, in Maharashtra (Rambaut et al., 2020). B.1.617.2 was later designated as a VOC in May 2021, following a significant surge in cases both regionally and globally, and was subsequently renamed the Delta variant (World Health Organization, 2021). Retrospective investigations revealed that the Delta VOC first emerged in India in mid-September 2020, although its transmission did not escalate until March 2021 (McCrone et al., 2022). The surge in Delta cases was accompanied by an alarming increase in local mortality and hospitalization rates due to its heightened transmissibility and immune evasion (Zhan et al., 2022). These advantageous features allowed Delta VOC to overtake the Alpha VOC as the dominant global lineage, fueling new outbreaks and resurgences worldwide despite the advancement of vaccine uptake (Liu and Rocklöv, 2021; Zhang et al., 2022).

In this study, we aim to characterize the genetic diversity and evolutionary trends of the Delta VOC in India. To achieve this, we conducted a multilayered analysis of the SARS-CoV-2 genome using archived whole-genome SARS-CoV-2 sequences from a public database. By collecting database-derived sequences at three distinct time periods surrounding the emergence of Delta VOC in India, we aim to identify specific mutation patterns that characterize the state of the viral population immediately before VOC emergence and reflect the succession of Delta VOC as the dominant SARS-CoV-2 lineage in early 2021. Furthermore, we seek to identify genomic regions of SARS-CoV-2 that underwent positive or negative selection to further improve our understanding of the SARS-CoV-2 evolutionary progression.

2. Materials and methods

2.1. Establishment of study groups and data collection

Full-length SARS-CoV-2 genomes derived from clinical specimens in India were sourced from the GISAID database for three study groups (A, B, and C) based on their initial sampling dates (Khare et al., 2021). The study groups and their corresponding time periods of interest included (1) Group A, which represents the earlier stage of the COVID-19 pandemic in India, when all reported cases were caused by the initial SARS-CoV-2 Wuhan strain and its early descendants; (2) Group B, which reflects the pre-Delta period, which was immediately before the suspected emergence of Delta VOC; and (3) Group C, which corresponds to the Delta-dominant period, when the Delta VOC was well established in the population (Figure 1) (India Today Web Desk, 2020; World Health Organization, 2020; Choudhary et al., 2021). The duration of the study periods was adjusted to ensure the sample size of each group was both acceptable and comparable to the other groups, as the number of positive COVID-19 cases per day (and thus the available genomes) varied greatly among the time periods of the three groups.

To ensure the reliability and completeness of the SARS-CoV-2 genome data used in our analysis, the GISAID filters for high-quality and complete genome sequences were applied while retrieving viral genomes for each group. The retrieved viral genomes were then viewed in Molecular Evolutionary Genetics Analysis (MEGA X), which revealed multiple genomes with indications of poor sequence quality (Kumar et al., 2018). These indicators included large unreported gaps (suggestive of amplicon dropout) or an excessive number of ambiguous bases, potentially arising from sequencing errors. Considering the potential interference of unreliable sequences in our mutation and selection investigations, these genomes were excluded from further analyses if the indicators at any given locus exceeded 1% of the

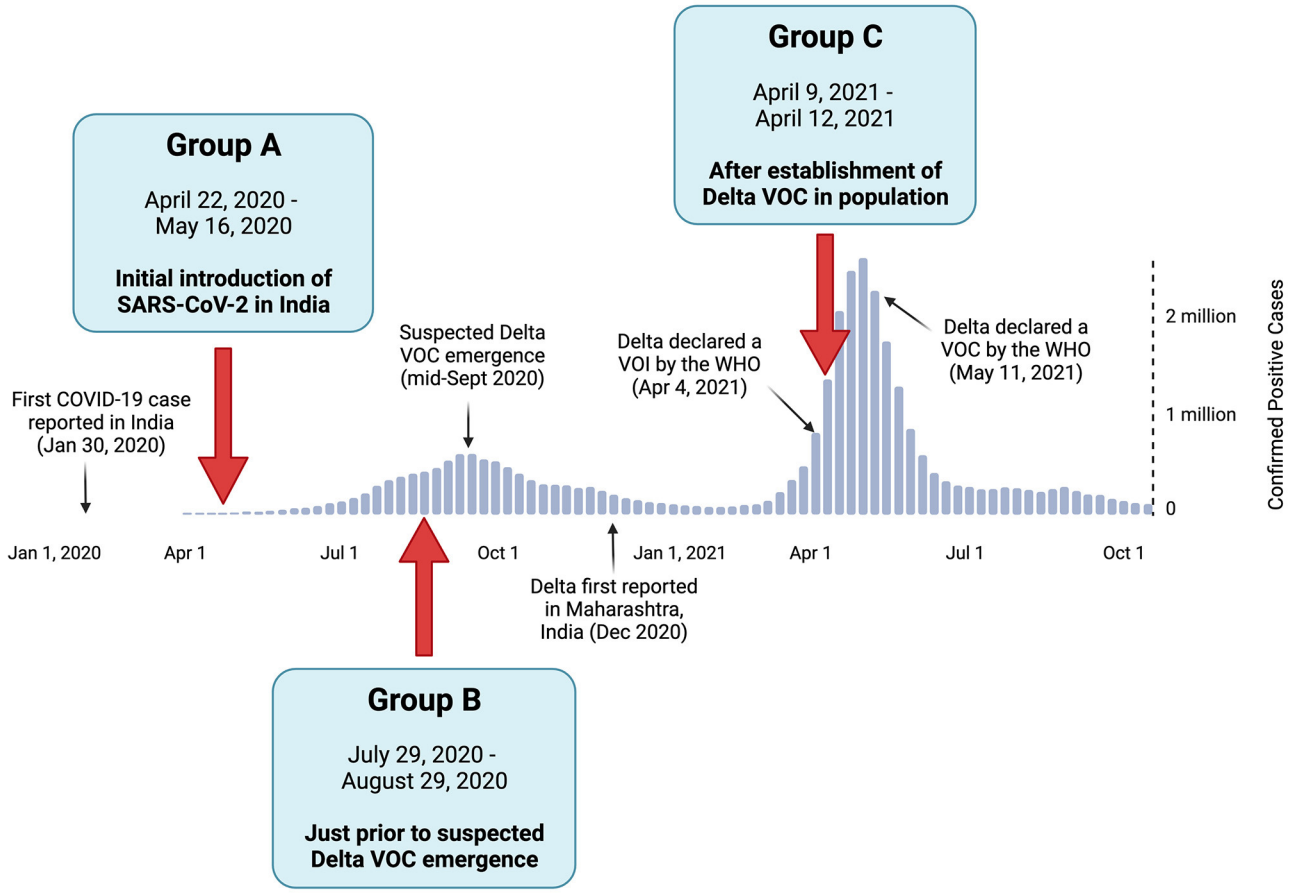


FIGURE 1
Positive COVID-19 case summary in India, with the relative timeline of events pertaining to the Delta variant of concern (VOC). The positive case numbers were based on the World Health Organization reports (World Health Organization, 2020). The collection dates and rationale of study Groups A, B, and C are denoted by red arrows.

total group's sequences. The detailed data retrieval and processing workflow is depicted in Figure 2.

2.2. Multiple sequence alignment

For each study group, all viral genome sequences were aligned to the original Wuhan-Hu-1 SARS-CoV-2 reference sequence (GISAID Accession EPI_ISL_402124) using the MAFFT v7 online server option for SARS-CoV-2 (https://mafft.cbrc.jp/alignment/server/add_sarscov2.html?mar15) (Katoh et al., 2019). The derived multiple sequence alignment (MSA) files were visually inspected in MEGA X (Kumar et al., 2018).

2.3. Mutation identification

Each group's full genome MSA was used to generate a positional nucleotide numeric summary file in BioEdit (Hall, 1999). This output file was then split into the individual genes of SARS-CoV-2 using an in-house Python program (https://github.com/connor-lowey/SARS-CoV-2_Delta_Helper_Scripts). To determine the prevalence of the mutations, the frequencies of all identified genetic variations were calculated and categorized using thresholds of <1%, 1–10%, and >10% for each respective group. The chosen thresholds stopped at >10% to accommodate heterogeneous viral populations and to avoid excluding an excess of data by setting an unattainable threshold. Mutations across the whole viral genome were characterized for all groups by recording mutations present at each frequency threshold in each individual gene relative to the SARS-CoV-2 reference genome. Potential codon changes at the amino acid level were recorded as resulting in a synonymous or non-synonymous amino acid substitution, an insertion, or a deletion. All mutations within a given gene at both the nucleotide and amino acid levels were compared among groups to determine which mutations were shared and which were unique to a particular group. The intergroup differences in mutation prevalence were further assessed for statistical significance using an in-house Python script (https://github.com/connor-lowey/SARS-CoV-2_Delta_Helper_Scripts). A chi-squared test of independence was used to compare groups by individual gene, and a Bonferroni correction was applied. A $p < 0.05$ was considered statistically significant.

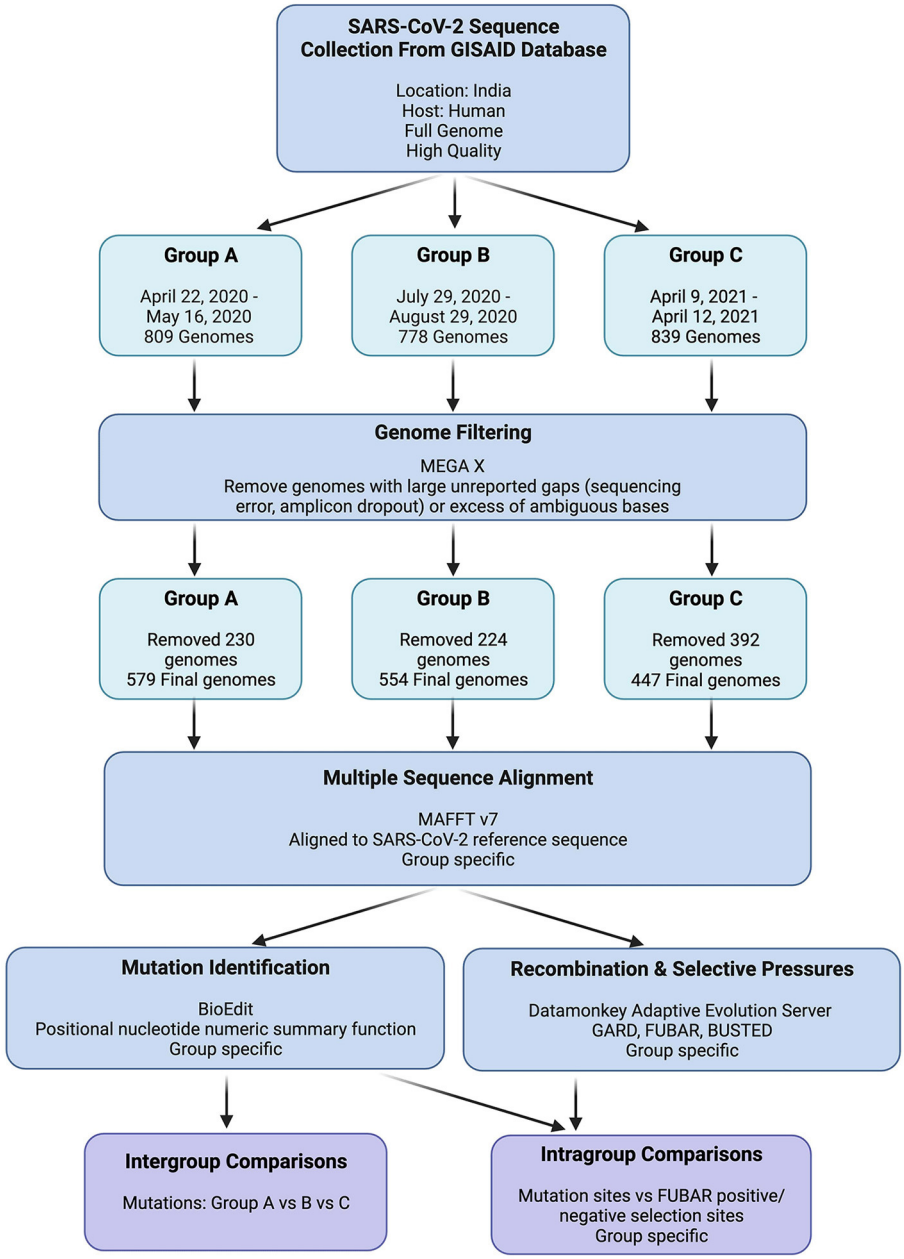


FIGURE 2
GISAID SARS-CoV-2 genome data retrieval and processing workflow. "Group Specific" indicates that individual analyses were performed on Groups A, B, and C independently.

2.4. Recombination and selection

To analyze recombination events and evidence of natural selection, we processed each group's whole-genome MSA file by splitting it into individual gene ORFs using MEGA X and then tested them for recombination and evidence of selection on the Datamonkey webserver (Kumar et al., 2018; Weaver et al., 2018). The recombination events were first assessed using the Genetic Algorithm for Recombination Detection (GARD) with site-to-site rate variation set to *General Discrete* and rate classes set to three (Kosakovsky Pond et al., 2006). All other parameters were left as default settings. Before conducting selection analyses, all stop codons (terminal and premature) were removed from the MSA files with MEGA X. This step was necessary as the Datamonkey selection tools reject the presence of any stop codons (Kumar et al., 2018). Gene-specific MSA files, omitting stop codons, were used to detect site-specific positive/negative selection using the Fast, Unconstrained Bayesian AppRoximation for Inferring Selection (FUBAR) tool with default parameter settings (Murrell et al., 2013). All amino acid sites under positive or negative selection were compared against the list of mutations in the corresponding group to determine which sites under selection coincided with a mutation

present at a frequency of >1% in the group. The intergroup differences in the quantities of positive and negative selection sites were assessed for statistical significance using an in-house Python script, as described in the section above.

3. Results

3.1. Establishment of study groups

The collection dates for each of the three study groups were selected based on the timeline of SARS-CoV-2 and the emergence of Delta VOC in India. As shown in Figure 1, Group A contained sequences collected from 22 April to 16 May 2020 ($n = 579$) during the initial introduction of the SARS-CoV-2 Wuhan strain to India. Group B was sampled from 29 July to 29 August 2020 ($n = 554$), representing the period just before the suspected emergence of the Delta VOC. Group C contained sequences from 9 April to 12 April 2021 ($n = 447$) during the surge in positive cases predominantly attributed to the Delta VOC. The most prevalent viral lineage found in Group A was B.1, which was the dominant global lineage during the start of the pandemic (Rambaut et al., 2020). In contrast, the most prevalent lineages were B.1.1.306 in Group B (a descendent lineage that did not achieve any notable WHO designations) and B.1.617.2 in Group C (the Delta VOC, as designated by the WHO). This highlights the diversification away from the original strain over time (Rambaut et al., 2020; World Health Organization, 2021).

3.2. Mutation analysis

To assess the temporal progression of mutation frequency and type across Groups A, B, and C, we first identified all mutations that are present at frequencies >1% in each of the study groups. Further mutation profiling revealed varying proportions of non-synonymous and synonymous codon changes among the groups, with Group C showing more distinctive results while Groups A and B were more alike. In all study groups, the ORF1a gene had the largest number of mutations out of any gene, as it makes up over 40% of the viral genome (Figure 3A) (Khare et al., 2021). Following closely behind in both Groups A and C were ORF1b and the S gene, as well as the N gene.

Group A was found to have the lowest number of nucleotide mutations (96), including the lowest quantity of both non-synonymous and synonymous substitutions out of all groups (Figure 4). However, when the relative percentage of each mutation type was calculated for the study groups, Group A had the largest proportion of non-synonymous mutations out of all the study groups. Specifically, this mutation type made up 61.5% of mutations in Group A and 60.5% and 51.7% of mutations in Groups B and C, respectively. Forty-eight (60.0%) of all mutations present at the 1–10% sequence threshold and 11 (68.8%) of all mutations present at the >10% sequence threshold encoded a non-synonymous substitution (Figure 4). Of these, the greatest number of non-synonymous substitutions in Group A occurred in ORF1a, with 17 at the 1–10% threshold range and 3 at the >10% threshold (Supplementary Table 1A). Notably, Group A had no insertions and only four deletions, all occurring in the E gene at the 1–10% threshold (Supplementary Table 1A).

Group B exhibited an intermediate quantity of mutations, with a higher total count than Group A but fewer mutations than Group C across all frequency ranges (Figures 3A–C). Out of the 114 mutations in Group B, 96 (84%) fell within the 1–10% frequency range, while only 18 mutations (16%) occurred in >10% of the group's sequences. This proportion of mutations reaching the >10% threshold was the lowest compared to the other groups (17% in Group A, 33% in Group C), although Group B had noticeably more mutations overall than Group A (114 vs. 96 mutations >1%, respectively). Most mutations were present in ORF1a and ORF1b, followed next in quantity by the N gene rather than the S gene (as seen in Groups A and C). Group B had the highest relative proportion of synonymous mutations across all frequency ranges (37.8% overall, Figure 4), most of which were found in ORF1a (Supplementary Table 1B). Similar to Group A, Group B did not contain any insertions and had the fewest deletions among all groups.

Group C had the highest mutation count at all frequency ranges among the three study groups (Figures 3A–C), with 2.2 and 1.85 times more mutations than the Group A and B totals at the >1% level, respectively. These differences were statistically significant, with $p<0.001$ for both pairwise comparisons of Groups A and B vs. Group C. Similarly, this group had significantly more mutations present at a frequency of >10%, with 4.4 and 3.9 times more than Groups A and B ($p < 0.001$), respectively. Interestingly, Group C also had the highest number of deletions present at all frequencies, and these deletions made up 22.7% of the group's total mutations shown in Figure 4. Consequently, the relative proportions of non-synonymous and synonymous substitutions were observed at lower amounts in Group C in comparison to the other groups; however, Group C had more mutations that cause a change in the amino acid coding (sum of non-synonymous substitutions, insertions, and deletions) than any other group at all frequencies (74% of all mutations present in >1% of sequences). Most of Group C's deletions were found in the ORF1a, ORF1b, S, and ORF8 genes (Supplementary Table 1C). Compared to Groups A and B, Group C contained significantly larger mutation counts in the S gene at all frequency levels ($p<001$). At a frequency of >10%, the S gene in Group C had one more mutation than ORF1a (21 and 20, respectively), despite ORF1a being 3.5 times longer than the S gene (Khare et al., 2021). Similarly, Group C also reported a significant difference in the number of mutations in ORF1a and ORF8 when compared to other study groups, as shown in Figure 3A ($p < 0.05$ and $p < 0.01$, respectively).

All mutations occurring in >1% of their respective group were compared to determine common and unique mutations within Groups A–C. Across the entire genome, a total of 42 amino acid coding sites were affected by mutations shared by at least two of the three study groups (Table 1). These involved 49 different nucleotide-level mutations, accounting for instances where multiple mutations occurred within the same codon. Notably, ORF7a, ORF7b, and ORF10 were the only genes that contained no shared mutations between study groups. On the other hand, ORF1a alone had almost half (48%) of the shared mutations in the entire genome, with 20 shared amino acid sites resulting

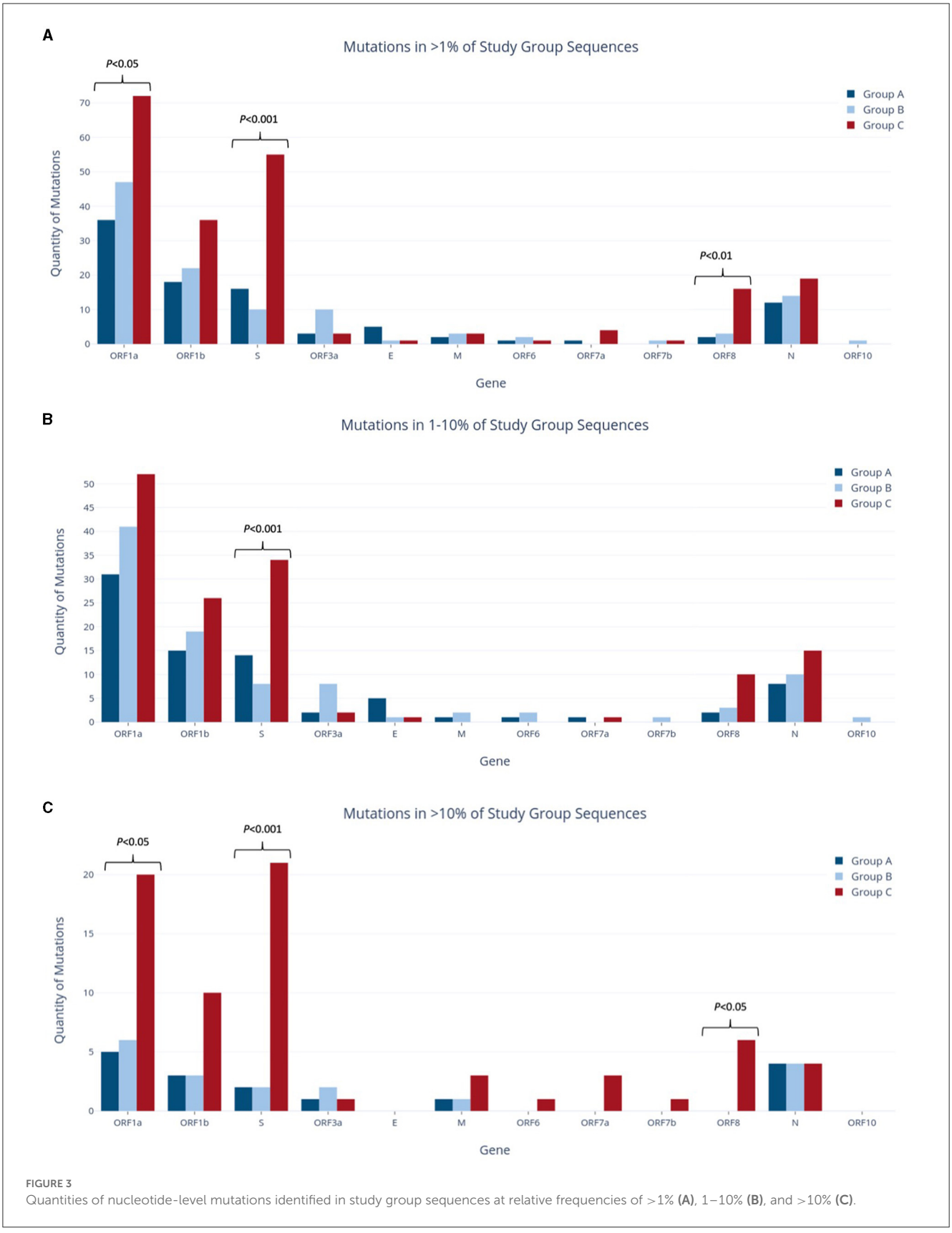


FIGURE 3
Quantities of nucleotide-level mutations identified in study group sequences at relative frequencies of >1% **(A)**, 1–10% **(B)**, and >10% **(C)**.

from 23 unique nucleotide-level mutations. ORF1b and N held the next-largest numbers of shared amino acid sites (seven and five, respectively), although these were considerably lower than the quantity of ORF1a. The most common combination for shared mutations was Group A and Group B, accounting for 25 of the 42 possible amino acid sites. Nine amino acid sites contained

10.3389/fmicb.2023.1222301

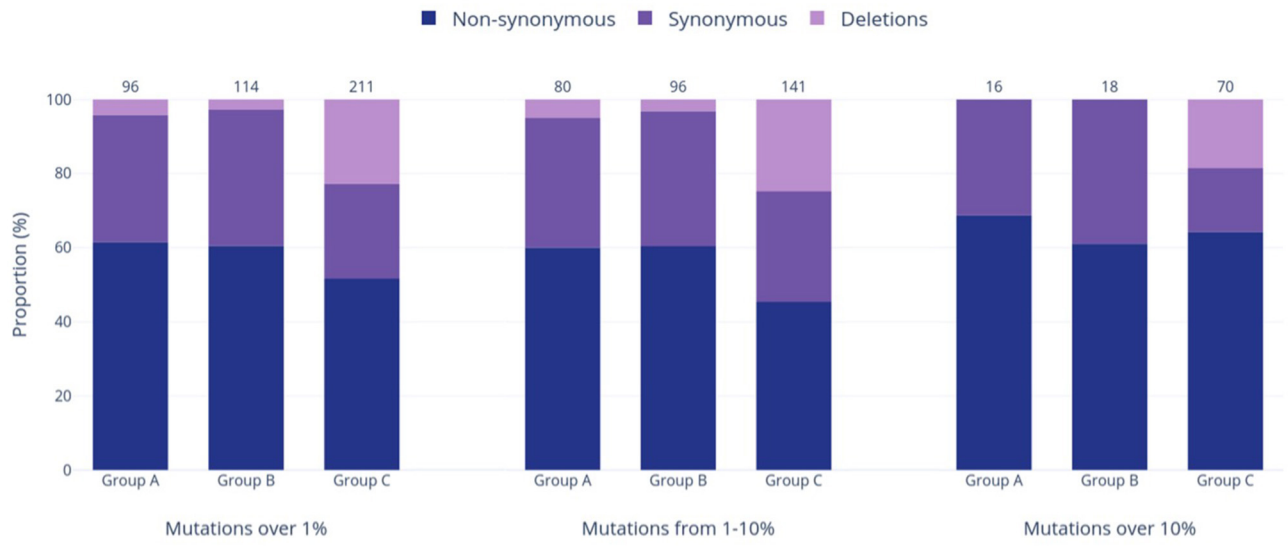


FIGURE 4
Prevalence of varied genetic mutation types in the three study groups at relative frequencies of >1%, 1–10%, and >10%. The bars in the graph represent the occurrence of each mutation type, while the numbers above the bars show the total counts of mutations within each respective group and frequency range.

TABLE 1 Quantities of mutated amino acid sites shared by multiple study groups.

Gene	Mutated amino acid sites shared by ≥2 study groups	Unique mutations[a] affecting shared amino acid sites
ORF1a	20	23
ORF1b	7	8
S	4	5
ORF3a	2	2
E	1	1
M	1	1
ORF6	1	1
ORF7a	0	0
ORF7b	0	0
ORF8	1	1
N	5	7
ORF10	0	0
Full genome	42	49

[a]Mutations, reflecting a single mutation at the nucleotide level where multiple distinct mutations (substitutions and/or deletions) can occur within the same amino acid codon.

mutations shared by Groups A, B, and C together, while six amino acid sites had mutations shared by just Groups B and C. The least frequent pairing was Group A with Group C, which only occurred at two amino acid sites in the full genome.

To quantify the unique mutations that were specific to each study group (not shared with another group), comparisons were made using genes, and the results are summarized in Table 2. Notably, unique mutations made up most of the mutations identified in Group C, at 91.5% (193 of the total 211 mutations present in >1% of Group C). The proportions of Group A and B were considerably lower than Group C, with unique mutations comprising 60.4% and 63.2% of all mutations in each group, respectively. Many of Group C's unique mutations were found in ORF1a, ORF1b, and the S gene. In ORF1a, Group C had 2.4- and 3.3 times more unique mutations than Groups B and A, respectively. Similarly, in ORF1b, Group C had 2.1- and 2.9 times more unique mutations than Groups B and A, respectively. The S gene of Group C exhibited a strikingly higher unique mutation count, with 4.3- and 7.4 times more mutations than those identified in Groups A and B, respectively.

3.3. Tests for recombination and selection

For each group, individual gene alignments were tested for recombination and then for the presence of positive or negative selection at individual amino acid sites to characterize the evolutionary context of each study group. As Datamonkey tools reject stop codons, the MSA input files were required to have all stop codons replaced with N's. This included the replacement of premature stop codons found in at least one sequence within the ORF1b, S, ORF3a, M, ORF6, ORF7a, ORF7a, ORF7b, ORF8, and N genes in various study groups. Analysis of gene alignments using GARD revealed no evidence of recombination within any individual gene across the study groups, so subsequent testing for positive selection could proceed without further adjustments. The findings from the FUBAR tool, as depicted in Figure 5, indicated that both positive (diversifying) and negative (purifying) selection was detected in all study groups, although the number

10.3389/fmicb.2023.1222301

TABLE 2 Unique nucleotide mutations occurring exclusively within individual study groups, with frequencies >1%.

Gene	Group A mutations	Proportion of total Group A mutations (%)[a]	Group B mutations	Proportion of total Group B mutations (%)[a]	Group C mutations	Proportion of total Group C mutations (%)[a]
ORF1a	20	55.6	27	57.4	65	90.3
ORF1b	11	61.1	15	68.2	32	88.9
S	12	75.0	7	70.0	52	94.5
ORF3a	1	33.3	8	80.0	3	100.0
E	4	80.0	0	0.0	1	100.0
M	1	50.0	2	66.7	3	100.0
ORF6	0	0.0	1	50.0	1	100.0
ORF7a	1	100.0	0	0.0	4	100.0
ORF7b	0	0.0	1	100.0	1	100.0
ORF8	2	100.0	2	100.0	15	93.8
N	6	50.0	8	57.1	16	84.2
ORF10	0	0.0	1	100.0	0	0.0
Full genome	58	60.4	72	63.2	193	91.5

[a] "Total mutations" refers to all genetic variations present within a group's genome sequences at frequencies exceeding 1%.

of amino acid sites reported varied greatly amongst the different genes.

Across the whole genome, the number of amino acid sites experiencing positive selection showed a progressive increase, from 33 sites in Group A to 47 in Group B and 51 in Group C, although no statistically significant difference was observed for any individual genes (Figure 5A). Although Group A had fewer sites under positive selection than Group B, it reported an equivalent number of genes affected. In contrast, Group A had the fewest genes experiencing negative selection, with only five genes vs. the nine and seven genes reported in Groups B and C, respectively. Of note, the ORF1a and S genes in Group A had the highest presence of negative selection, at 16 and 11 sites, respectively. Similarly, ORF1a and ORF1b had the most sites under positive selection, with 8 and 10 sites, respectively. Interestingly, the S gene in Group A had 2.75 times more sites under negative selection than positive selection.

Group B reported a significant increase in the number of sites under negative selection, with 99 sites identified across the whole genome. This represented a substantial difference as compared to Group A, which had 38 sites (2.6 times fewer, $p < 0.001$), and Group C, which had 41 sites (2.4 times fewer, $p < 0.001$). As shown in Figure 5B, most of these sites were found in ORF1a and ORF1b, with 45 and 34 sites, respectively. The difference between study groups was highly significant for both of these genes ($p < 0.001$). As stated above, Group B also had the highest number of genes affected by negative selection out of all groups. ORF1a and ORF1b also contributed the largest number of sites under positive selection in Group B, although on a much smaller scale than those under negative selection (16 and 8 sites, respectively). In contrast to Group A, the S gene in Group B contained more sites under positive selection (seven sites) but fewer sites under negative selection (five sites). Additionally, Group B reported seven sites under positive selection in ORF3a, while Group A reported two sites and Group C reported only one in this gene. This difference in ORF3a was initially statistically significant ($p < 0.05$) but was no longer significant after applying the Bonferroni correction.

Group C had the highest number of amino acid sites under positive selection, with a total of 51 sites across the entire genome. Importantly, Group C was the only group in which the number of sites under positive selection exceeded those under negative selection, as both Groups A and B reported the opposite pattern at the genome level. Group C also found nine genes with sites experiencing positive selection, compared to seven genes in both Groups A and B. ORF1a had an equivalent number of sites under both positive and negative selection (18), while ORF1b was nearly equivalent with 11 sites under positive selection and 10 under negative selection. The S gene in Group C also had a higher number of sites under positive selection (11 sites) as compared to negative selection (six sites), with four of these positive selection sites corresponding to lineage-defining mutations of the Delta VOC (G142D, L452R, P681R, and D950N; Table 3).

To assess the relationship between the sites under positive or negative selection and mutations present at frequencies >1%, comparisons were made within each study group at the amino acid level. The proportion of sites under selection that also had a mutation at the same amino acid site is summarized by the gene in Supplementary Tables 2A–C. As the number of sites under selection varied among genes (Figure 5), numbers were presented as the percentage of sites under selection for ease of comparison. It should be noted that this study focused exclusively on mutations present in >1% of each study group, and consequently, not all sites under selection had corresponding mutations at the same locus that met the >1% threshold. As a result, the overall proportion of selection sites with matching mutations was low across the full genome for all three study groups, although certain genes exhibited higher proportions (Supplementary Tables 2A–C).

FIGURE 5
Quantities of amino acid sites identified as experiencing positive selection **(A)** or negative selection **(B)** in study groups.

Importantly, Group C contained all the expected lineage-defining mutations of the Delta VOC, as described in Table 3 (cov-lineages.org, 2023; Hodcroft, 2023d). Several mutations, including F924F in ORF1a, P314L in ORF1b, and D614G in S, were identified in all study groups, supporting the establishment of these mutations in the viral population at an early stage. In particular, the P314L mutation in ORF1b, which was under positive selection in all three study groups, and the D614G Mutation in S are also lineage-defining mutations for the Alpha, Beta, and Gamma VOCs, all of which were detected earlier than the Delta VOC (Hodcroft, 2023a,b,c,d). The S protein position 681 holds a lineage-defining mutation for both the Alpha and Delta VOCs, although the amino acid substitutions differ between lineages, and was reported to be under positive selection in Group C (Hodcroft, 2023b,d). Similarly, the N protein position 203 serves as a lineage-defining mutation site for the Alpha, Gamma, and Delta VOCs, although the Alpha and Gamma lineages share the R203K substitution while Delta contains R203M (Hodcroft, 2023b,c,d). Additionally, both Group A and C reported the G662S mutation in ORF1b, but this is only a lineage-defining mutation of the Delta and Omicron BA.2.75, XBB, and XBB.1.5 VOCs (equivalent Nextstrain nomenclature are Omicron 22D, 22F, and 23A, respectively), all of which emerged long after the sequence collection dates of Group A (Hodcroft, 2023d,f,g,h).

4. Discussion

The error-prone nature of RNA viruses and the massive global availability of susceptible hosts have provided ample opportunity for SARS-CoV-2 to circulate and diversify rapidly (Duarte et al., 2022). Following the succession of the Delta VOC (B.1.617.2) as the dominant global variant, continued adaptive evolution led to the emergence of multiple sub-lineages within the Delta variant family, termed AY lineages by Pango nomenclature (Eales et al., 2022; cov-lineages.org, 2023). Here, we focused on the viral population in India to characterize trends in genetic diversity and selection before

10.3389/fmicb.2023.1222301

TABLE 3 Delta VOC lineage-defining mutations identified in study Groups A–C.

Gene	Mutation	Identified groups	Selection	References
ORF1a	F924F	A, B, C	Negative (A, B)	[a]
ORF1b	P314L	A, B, C	Positive (A, B, C)	[a]
ORF1b	G662S	A, C	None	[a]
ORF1b	P1000L	C	None	[a]
S	T19R	C	None	[a,b]
S	G142D	C	Positive	[a]
S	E156-	C	None	[a]
S	F157-	C	None	[a]
S	R158G	C	None	[a]
S	L452R	C	Positive	[a,b]
S	T478K	C	None	[a,b]
S	D614G	A, B, C	None	[a]
S	P681R	C	Positive	[a,b]
S	D950N	C	Positive	[a,b]
ORF3a	S26L	C	Positive	[a,b]
M	I82T	C	Positive	[a,b]
ORF7a	V82A	C	None	[a,b]
ORF7a	T120I	C	Positive	[a,b]
ORF8	D119-	C	None	[a]
ORF8	F120-	C	None	[a]
N	D63G	C	Positive	[a,b]
N	R203M	C	None	[a,b]
N	D377Y	C	None	[a,b]

[a]Hodcroft, 2023d at https://covariants.org/variants/21A.Delta.
[b]cov-lineages.org, 2023 at https://cov-lineages.org/global_report_B.1.617.2.html.

and after the emergence of Delta VOC. The range of collection dates for Groups A, B, and C provide distinct snapshots of the viral population at the first introduction of SARS-CoV-2 to India, just before the emergence of the Delta variant, and well after the Delta VOC was established in the country. This unique sampling strategy effectively displays the evolutionary progression toward VOC emergence. To the best of our knowledge, this is currently the only study providing an evolutionary perspective through this tri-phased sampling strategy.

As expected, mutation analyses revealed increasing quantities of mutations with the temporal progression of Groups A–C. Notably, Group C had a significant increase in mutations at both the >1% and >10% thresholds, which is well supported by reports showing that the Delta VOC harbors 29 characteristic mutations that differ from the original Wuhan strain (Khare et al., 2021; Borcard et al., 2022; cov-lineages.org, 2023; Gangavarapu et al., 2023). In particular, 33.2% of the total mutations found in Group C were present in >10% of the group's sequences, as opposed to 16.7% and 15.8% in Groups A and B, respectively. While the increasing quantity of mutations demonstrates an increase in genetic diversity, the larger percentage of mutations reaching the >10% threshold demonstrates the progression toward a new lineage becoming dominant in the viral population. Similarly, intergroup comparisons supported the evolutionary context of Groups A, B, and C. Comparisons of mutations revealed that mutations identified in multiple groups were most often shared between Groups A and B, whereas mutations shared between Groups A and C were rare. This is reasonable given that the collection dates of the groups progressed temporally from Group A to B to C, with Group A and C being collected almost a full year apart. Furthermore, 91.5% of mutations found in >1% of Group C's sequences were unique mutations not shared by any other group, which emphasizes the distinct genetic profile that appears after the emergence of Delta VOC. Despite having lower proportions than Group C, Groups A and B still reported over 60% of their mutations as unique rather than shared, demonstrating that the viral population of each group possesses a specific genetic profile that changes over time.

Another interesting pattern in Group C was the proportion of different mutation types, which differed from Groups A and B. At all examined frequency thresholds (Figure 4), all three groups had a higher percentage of non-synonymous substitutions than synonymous substitutions, but Groups A and B consistently reported similar percentages in each mutation type. While Group C still followed this pattern, the percentages of non-synonymous and synonymous substitutions were distinct, and the proportion of synonymous mutations was lower than those of Groups A and B at all frequencies. This was exchanged for a large increase in deletions, with Group C reporting 12–16 times more deletions than Groups A and B overall (>1% threshold) and additionally being the only group that contained deletions (13 occurrences) above the >10% sequence threshold. Specifically, deletions within the S gene made up 22 of the 48 identified Group C deletions, which is not surprising as S protein deletions have been reported to affect transmissibility, antigenicity, and immune escape, thus conferring a fitness advantage (Harvey et al., 2021; Liu et al., 2021). Deletions have been frequently reported throughout the genome of all five VOC lineages, including the Delta VOC, which dominates Group C (Hodcroft, 2023a,b,c,d,e). No insertions were reported in >1% of sequences in any group, which was unsurprising given that insertions are known to occur far less frequently than deletions in both the SARS-CoV-2 virus and in broader protein evolution (De Jong and Rydén, 1981; Liu et al., 2021). Altogether, Group C has the largest proportion of mutations causing changes in the final amino acid code, such as non-synonymous substitutions and deletions. In total, 74.4% of mutations in Group C at frequencies >1% caused changes in the final amino acid code (vs. 65.6 and 63.2% in Groups A and B, respectively). This proportion increased to 82.9% of mutations at the >10% threshold (vs. 68.8 and 61.1% in Groups A and B, respectively). This tendency to cause a change in the final genome sequence supports the dynamic nature of the SARS-CoV-2 virus and explains its incredible ability to adapt to human populations by continuously diversifying and evolving. Similarly, other RNA viruses such as human immunodeficiency virus 1 (HIV-1) are known to be highly mutable viruses as well due to the activity of the error-prone RNA-dependent RNA polymerase,

which creates ample opportunity for diversification (Kustin and Stern, 2021; Duarte et al., 2022).

Certain mutations may become established because of natural selection as favorable changes can provide evolutionary advantages such as improved survival and reproduction of the virus (Karlsson et al., 2014). Shortly after SARS-CoV-2 emergence, selection favored mutations that contributed to reproductive success, such as the S gene D614G substitution, as there was little evolutionary need for antigenic diversity (Carabelli et al., 2023). Interestingly, the D614G mutation was not reported to be under any selection, despite being detected in all study groups (Table 3). This is perhaps because it became an established mutation very early in the pandemic (first reported in January 2020) and was highly prevalent in all our study groups, potentially causing it to be regarded as the standard sequence rather than a mutation (Tian et al., 2021).

As natural and vaccine-acquired immunity against SARS-CoV-2 increased in the host population, so did the need for viral antigenic variations to enable its immune escape and continue transmission (Carabelli et al., 2023). Mutations in viral surface proteins highly exposed to the immune system, such as the S protein, are typically under high selective pressure, and the Delta VOC holds the majority of its characteristic mutations within the S protein (Malik et al., 2021; cov-lineages.org, 2023). For example, the S protein L452R mutation was identified in Group C and found to be under positive selection in our study. Located in the receptor-binding motif region of the RBD, which binds the host angiotensin-converting enzyme-2 (ACE2) receptor, L452R causes structural changes that may stabilize the interaction between the S protein and the ACE2 receptor on the host cell to increase viral infectivity (Tian et al., 2021). Similarly, the S protein P681R mutation was identified in Group C and reported to be experiencing positive selection. This mutation sits within the furin cleavage site of the S protein, and this cleavage of the S1 and S2 subunits is a critical part of host cell entry (Huang et al., 2020; Tian et al., 2021). The P681R mutation in the S protein facilitates furin-mediated cleavage, which improves host cell entry, and was reported to be an important element of succession of Delta VOC over the previously dominant Alpha VOC (Tian et al., 2021; Liu et al., 2022). This list is not exhaustive but conveys the importance of these high-prevalence mutations found in Group C and supports how selection drives the evolution of viral lineages with an advantageous repertoire of mutations. Evidence of both positive and negative selection has been reported by other studies in the S gene, while positive selection has been reported in ORF1ab, ORF3a, and ORF8 (Velazquez-Salinas et al., 2020; Martin et al., 2021; Duarte et al., 2022; Upadhyay et al., 2022).

The number of amino acid sites experiencing positive selection increased with the temporal group progression, from 33 in Group A to 47 in Group B and 51 in Group C. Most of these sites were in ORF1a and ORF1b, which is unsurprising given that they encompass around 70% of the viral genome together (Bai et al., 2022). Other genes reporting high numbers of variations in all groups included S and N, while the remaining structural genes (E and M) found hardly any sites under positive selection. Interestingly, Group B was the only group to report that ORF3a (functions in viral release, inflammasome activation, and necrotic cell death) had a higher number of sites under positive selection, equivalent to the S and N genes (Naqvi et al., 2020; Gorkhali et al., 2021; Bai et al., 2022). In terms of negative selection, Group B had 2.4–2.6 times more sites detected within the whole genome than any other group, with ORF1a and ORF1b contributing a large number of sites. ORF1b notably reported a jump from 5 sites in Group A to 10 sites in Group C and 34 sites in Group B. Additionally, Group B contains more mutations reaching the >1% threshold than Group A, but fewer mutations reaching the >10% threshold. Given that Group B represents an intermediate period where SARS-CoV-2 is established and circulating in the country, but the Delta VOC has not yet emerged, we suspect that these results demonstrate a period of heightened diversification within the viral population.

As the virus moves away from its initial introduction in India (Group A), it gains more opportunities to diversify by increasing the number of infected hosts and by the progression of time, allowing more viral replication and transmission cycles (Carabelli et al., 2023). This heightened genetic diversity is evident in Group B as characterized by an increased total mutation count; however, many of these did not reach the >10% threshold, which indicates the presence of numerous low-prevalence lineages rather than a single dominant lineage. Similarly, Group B reported the highest proportion of synonymous mutations, which do not change the translated amino acid code. This observation aligns with the findings that Group B had the largest number of sites experiencing negative selection across all study groups. During viral evolution, negative selection plays a crucial role in preserving essential functional features through synonymous substitutions (Spielman et al., 2019; Berrio et al., 2020). As RNA viruses are known to experience strong purifying selection due to their densely coded genome, it is plausible that the observed increase in Group B's negative selection is working to maintain essential gene function while the viral population is undergoing a massive expansion and diversifying into new lineages (Kustin and Stern, 2021). This notion is supported by the work of Martin et al., who reported a significant shift in selective pressures within the global SARS-CoV-2 viral population about 11 months after its initial appearance (November 2020), where the number of sites detected under both positive and negative selection increased substantially and continued throughout the following months in 2021. Remarkably, the time frame in the study by Martin et al. coincides with the collection dates of Groups B and C in our work, further supporting the differentiation of Groups B and C from Group A with respect to the quantities of sites under selection.

One notable limitation of this study is its dependency on the GISAID database for viral genome availability and quality. The detection of viral variants heavily depends on the extent of local sequencing efforts, the throughput capacity of Indian sequencing laboratories, and the consistent deposition of Indian viral sequences in the GISAID database. We encountered challenges during the study of Group A, which was sampled during the early stage of the COVID-19 pandemic in India when viral genome availability in GISAID was very limited. Consequently, both Groups A and B were sampled over much wider collection date ranges than Group C to provide a similar number of genomes across all groups. This suggests a less precise snapshot of genetic diversity as the wider ranges of collection dates may include multiple cycles of viral transmission. Similarly, our study relied on the

availability of high-quality sequences within databases. Despite applying quality filters during database queries, many samples were removed from our study due to the presence of unreliable genome regions suspected of poor-quality sequencing and/or amplicon dropout. This was especially troublesome in Group C, where the increased prevalence of mutations might have made sequencing with previous primer schemes increasingly difficult, which was a known problem for the Delta VOC (Borcard et al., 2022). According to the periods of sequence collection, Group A may have used ARTIC primer V2 or earlier, while Group B and C may have used ARTIC primer V3 or earlier (Quick, 2020). The ARTIC primer V4 series, designed to address SARS-CoV-2 Beta, Gamma, and Delta VOC mutations in V3 primer binding sites, was released on June 18, 2021, which was several months after the collection dates of Group C (Davis et al., 2021). Nonetheless, we endeavored to address these limitations through the utilization of a larger number of genomes within each group, meticulous attention to genome quality, and a focus on whole-genome analysis.

This study effectively demonstrates the evolutionary progression toward VOC emergence in SARS-CoV-2, with a specific focus on the Delta VOC, known for its significant global impact as one of the most virulent variant lineages. With a unique tri-phased sampling strategy, this study provides valuable insight into the evolutionary dynamics preceding and following the emergence of Delta VOC. It is plausible that similar patterns of genetic diversity and evolutionary selection may have been observed in other VOC lineages before their initial appearance. By enhancing our understanding of the dynamic nature of SARS-CoV-2 evolution, this study has the potential to facilitate earlier recognition and prediction of the emergence of future variant lineages. Such insights could be instrumental in mitigating the impact of the emerging variants and effectively responding to the evolving challenges from the SARS-CoV-2 VOCs.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

KL: conceptualization and formal analysis. KL and SM: methodology and writing—original draft preparation. KL, SM, PS, and HJ: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This study was partially supported by the operating budget of the National Microbiology Laboratory, Public Health Agency of Canada, to which the authors are affiliated. KL is also a recipient of a research scholarship from Research Manitoba, Canada and the Visual and Automated Disease Analytics (VADA) Graduate Training Program, Canada.

Acknowledgments

The authors gratefully acknowledge all data contributors, including the authors and their originating laboratories for providing the specimens, as well as their submitting laboratories for generating the genetic sequence and metadata and sharing *via* the GISAID Initiative, on which this study is based. Figures 1, 2 were created with BioRender. The authors would also like to thank Connor Lowey for writing the in-house programs used in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1222301/full#supplementary-material

References

Andrews, M. A., Areekal, B., Rajesh, K. R., Krishnan, J., Suryakala, R., Krishnan, B., et al. (2020). First confirmed case of COVID-19 infection in India: a case report. *Indian J. Med. Res.* 151, 490–492. doi: 10.4103/ijmr.IJMR_2131_20

Bai, C., Zhong, Q., and Gao, G. F. (2022). Overview of SARS-CoV-2 genome-encoded proteins. *Sci. China Life Sci.* 65, 280. doi: 10.1007/s11427-021-1964-4

Berrio, A., Gartner, V., and Wray, G. A. (2020). Positive selection within the genomes of SARS-CoV-2 and other Coronaviruses independent of impact on protein function. *PeerJ.* 8, e10234. doi: 10.7717/peerj.10234

Borcard, L., Gempeler, S., Terrazos Miani, M. A., Baumann, C., Grädel, C., Dijkman, R., et al. (2022). Investigating the extent of primer dropout in SARS-CoV-

2 genome sequences during the early circulation of delta variants. *Front. Virol.* 2, 28. doi: 10.3389/fviro.2022.840952

Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., de Silva, T. I., et al. (2023). SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* 1. doi: 10.1038/s41579-022-00841-7

Choudhary, O. P., Priyanka Singh, I., and Rodriguez-Morales, A. J. (2021). Second wave of COVID-19 in India: Dissection of the causes and lessons learnt. *Travel Med. Infect. Dis.* 43, 102126. doi: 10.1016/j.tmaid.2021.102126

cov-lineages.org (2023). *B.1.617.2 Report*. Available at: https://cov-lineages.org/global_report_B.1.617.2.html (accessed May 2, 2023).

Davis, J. J., Long, S. W., Christensen, P. A., Olsen, R. J., Olson, R., Shukla, M., et al. (2021). Analysis of the ARTIC version 3 and version 4 SARS-CoV-2 primers and their impact on the detection of the G142D amino acid substitution in the spike protein. *Microbiol Spectr.* 9, 21. doi: 10.1128/Spectrum.01803-21

De Jong, W. W., and Rydén, L. (1981). Causes of more frequent deletions than insertions in mutations and protein evolution. *Nature* 290, 157–159. doi: 10.1038/290157a0

Duarte, C. M., Ketcheson, D. I., Eguíluz, V. M., Agust,í, S., Fernández-Gracia, J., Jamil, T., et al. (2022). Rapid evolution of SARS-CoV-2 challenges human defenses. *Sci. Rep.* 12, 10097. doi: 10.1038/s41598-022-10097-z

Eales, O., Page, A. J., de Oliveira Martins, L., Wang, H., Bodinier, B., Haw, D., et al. (2022). SARS-CoV-2 lineage dynamics in England from September to November 2021: high diversity of Delta sub-lineages and increased transmissibility of AY.4.2. *BMC. Infect. Dis.* 22, 4. doi: 10.1186/s12879-022-07628-4

Gangavarapu, K., Latif, A. A., Mullen, J. L., Alkuzweny, M., Hufbauer, E., Tsueng, G., et al. (2023). Outbreak.info genomic reports: scalable and dynamic surveillance of SARS-CoV-2 variants and mutations. *Nat. Methods* 20, 512–522. doi: 10.1038/s41592-023-01769-3

Gorkhali, R., Koirala, P., Rijal, S., Mainali, A., Baral, A., and Bhattarai, H. K. (2021). Structure and function of major SARS-CoV-2 and SARS-CoV proteins. *Bioinform. Biol. Insights* 15. doi: 10.1177/11779322211025876

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., et al. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424. doi: 10.1038/s41579-021-00573-0

Hodcroft, E. B. (2023a). "Variant: 20H (Beta, V2)," in *CoVariants*. Available online at: https://covariants.org/variants/20H.Beta.V2 (accessed May 2, 2023).

Hodcroft, E. B. (2023b). "Variant: 20I (Alpha, V1)," in *CoVariants*. Available online at: https://covariants.org/variants/20I.Alpha.V1 (accessed May 2, 2023).

Hodcroft, E. B. (2023c). "Variant: 20J (Gamma, V3)," in *CoVariants*. Available online at: https://covariants.org/variants/20J.Gamma.V3 (accessed May 2, 2023).

Hodcroft, E. B. (2023d). "Variant: 21A (Delta)," in *CoVariants*. Available online at: https://covariants.org/variants/21A.Delta (accessed May 2, 2023).

Hodcroft, E. B. (2023e). "Variant: 21K (Omicron)," in *CoVariants*. Available online at: https://covariants.org/variants/21K.Omicron (accessed May 2, 2023).

Hodcroft, E. B. (2023f). "Variant: 22D (Omicron)," in *CoVariants*. Available online at: https://covariants.org/variants/22D.Omicron (accessed May 2, 2023).

Hodcroft, E. B. (2023g). "Variant: 22F (Omicron)," in *CoVariants*. Available online at: https://covariants.org/variants/22F.Omicron (accessed May 2, 2023).

Hodcroft, E. B. (2023h). "Variant: 23A (Omicron)," in *CoVariants*. Available online at: https://covariants.org/variants/23A.Omicron (accessed May 2, 2023).

Huang, Y., Yang, C., Xu, X., feng, Xu, W., and Liu, S., wen (2020). Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41, 1141–1149. doi: 10.1038/s41401-020-0485-4

India Today Web Desk (2020). Two with travel history to Dubai test positive for coronavirus in Pune. *India Today*. Available online at: https://www.indiatoday.in/india/story/two-with-travel-history-to-dubai-test-positive-for-coronavirus-in-pune-1654113-2020-03-09 (accessed May 2 2023).

Karamitros, T., Papadopoulou, G., Bousali, M., Mexias, A., Tsiodras, S., and Mentis, A. (2020). SARS-CoV-2 exhibits intra-host genomic plasticity and low-frequency polymorphic quasispecies. *J. Clini. Virol.* 131, 104585. doi: 10.1016/j.jcv.2020.104585

Karlsson, E. K., Kwiatkowski, D. P., and Sabeti, P. C. (2014). Natural selection and infectious disease in human populations. *Nat. Rev. Genet.* 15, 379–394. doi: 10.1038/nrg3734

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* 20, 1160–1166. doi: 10.1093/bib/bbx108

Khailany, R. A., Safdar, M., and Ozaslan, M. (2020). Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* 19, 100682. doi: 10.1016/j.genrep.2020.100682

Khare, S., Gurry, C., Freitas, L., Schultz, M. B., Bach, G., Diallo, A., et al. (2021). GISAID's role in pandemic response. *China CDC Weekly*, 2021. 3, 1049–1051. doi: 10.46234/ccdcw2021.255

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., and Frost, S. D. W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* 23, 1891–1901. doi: 10.1093/molbev/msl051

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kustin, T., and Stern, A. (2021). Biased mutation and selection in RNA viruses. *Mol. Biol. Evol.* 38, 575–588. doi: 10.1093/molbev/msaa247

Letko, M., Marzi, A., and Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat. Microbiol.* 5, 562–569. doi: 10.1038/s41564-020-0688-y

Li, M.-Y., Li, L., Zhang, Y., and Wang, X.-S. (2020). Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infect. Dis. Pov.* 9, 1–7. doi: 10.1186/s40249-020-00662-x

Liu, X., Guo, L., Xu, T., Lu, X., Ma, M., Sheng, W., et al. (2021). A comprehensive evolutionary and epidemiological characterization of insertion and deletion mutations in SARS-CoV-2 genomes. *Virus Evol.* 7, 1–10. doi: 10.1093/ve/veab104

Liu, Y., Liu, J., Johnson, B. A., Xia, H., Ku, Z., Schindewolf, C., et al. (2022). Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. *Cell Rep.* 39, 110829. doi: 10.1016/j.celrep.2022.110829

Liu, Y., and Rocklöv, J. (2021). The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *J. Travel Med.* 28. doi: 10.1093/jtm/taab124

Malik, Y. S., Ansari, M. I., Kattoor, J. J., Kaushik, R., Sircar, S., Subbaiyan, A., et al. (2021). Evolutionary and codon usage preference insights into spike glycoprotein of SARS-CoV-2. *Brief. Bioinformatics* 22, 1006–1022. doi: 10.1093/bib/bbaa383

Martin, D. P., Weaver, S., Tegally, H., San, J. E., Shank, S. D., Wilkinson, E., et al. (2021). The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell* 184, 5189–5200. doi: 10.1016/j.cell.2021.09.003

McCrone, J. T., Hill, V., Bajaj, S., Pena, R. E., Lambert, B. C., Inward, R., et al. (2022). Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature* 610, 154–160. doi: 10.1038/s41586-022-05200-3

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., et al. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.* 30, 1196–1205. doi: 10.1093/molbev/mst030

Naqvi, A. A. T., Fatima, K., Mohammad, T., Fatima, U., Singh, I. K., Singh, A., et al. (2020). Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochim. Biophys. Acta Mol. Basis Dis.* 1866, 165878. doi: 10.1016/j.bbadis.2020.165878

Quick, J. (2020). *nCoV-2019 sequencing protocol v3 (LoCost) V.3*. ARTIC Network. Available online at: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v3-locost-bp2l6n26rgqe/v3/materials.

Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., et al. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol.* 5, 1403–1407. doi: 10.1038/s41564-020-0770-5

Siddiqui, A. F., Wiederkehr, M., Rozanova, L., and Flahault, A. (2020). Situation of India in the COVID-19 pandemic: India's initial pandemic experience. *Int. J. Environ. Res. Public Health* 17, 1–18. doi: 10.3390/ijerph17238994

Spielman, S. J., Weaver, S., Shank, S. D., Magalis, B. R., Li, M., and Kosakovsky Pond, S. L. (2019). "Evolution of viral genomes: Interplay between selection, recombination, and other forces," in *Evolutionary Genomics: Statistical and Computational Methods, Methods in Molecular Biology* (Totowa, NJ: Humana Press Inc.), 427–468. doi: 10.1007/978-1-4939-9074-0_14

Tian, D., Sun, Y., Zhou, J., and Ye, Q. (2021). The global epidemic of the SARS-CoV-2 Delta variant, key spike mutations and immune escape. *Front. Immunol.* 12, 5001. doi: 10.3389/fimmu.2021.751778

Toft-Bertelsen, T. L., Jeppesen, M. G., Tzortzini, E., Xue, K., Giller, K., Becker, S., et al. (2021). Amantadine inhibits known and novel ion channels encoded by SARS-CoV-2 *in vitro*. *Communicat. Biol.* 4, 1–10. doi: 10.1038/s42003-021-02866-9

Upadhyay, V., Patrick, C., Lucas, A., and Mallela, K. M. G. (2022). Convergent evolution of multiple mutations improves the viral fitness of SARS-CoV-2 variants by balancing positive and negative selection. *Biochemistry* 61, 963–980. doi: 10.1021/acs.biochem.2c00132

Velazquez-Salinas, L., Zarate, S., Eberl, S., Gladue, D. P., Novella, I., and Borca, M. V. (2020). Positive selection of ORF1ab, ORF3a, and ORF8 genes drives the early evolutionary trends of SARS-CoV-2 during the 2020 COVID-19 pandemic. *Front. Microbiol.* 11, 550674. doi: 10.3389/fmicb.2020.550674

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., Kosakovsky Pond, S. L., et al. (2018). Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. *Mol. Biol. Evol.* 35, 773–777. doi: 10.1093/molbev/msx335

World Health Organization (2020). "WHO COVID-19 Dashboard," in *WHO Health Emergency Dashboard*. Available online at: https://covid19.who.int/ (accessed May 2, 2023).

World Health Organization (2021). *Tracking SARS-CoV-2 Variants*. Available at: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed January 17, 2022).

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3

Zhan, Y., Yin, H., and Yin, J.-Y. (2022). B.1.617.2 (Delta) Variant of SARS-CoV-2: features, transmission and potential strategies. *Int. J. Biol. Sci.* 18, 1844. doi: 10.7150/ijbs.66881

Zhang, X., Chen, S., Cao, Z., Yao, Y., Yu, J., Zhou, J., et al. (2022). Increased pathogenicity and aerosol transmission for one SARS-CoV-2 B.1.617.2 Delta variant over the wild-type strain in hamsters. *Virol. Sin.* 37, 796–803. doi: 10.1016/j.virs.2022.09.008

Zinzula, L. (2021). Lost in deletion: the enigmatic ORF8 protein of SARS-CoV-2. *Biochem. Biophys. Res. Commun.* 538, 116–124. doi: 10.1016/j.bbrc.2020.10.045

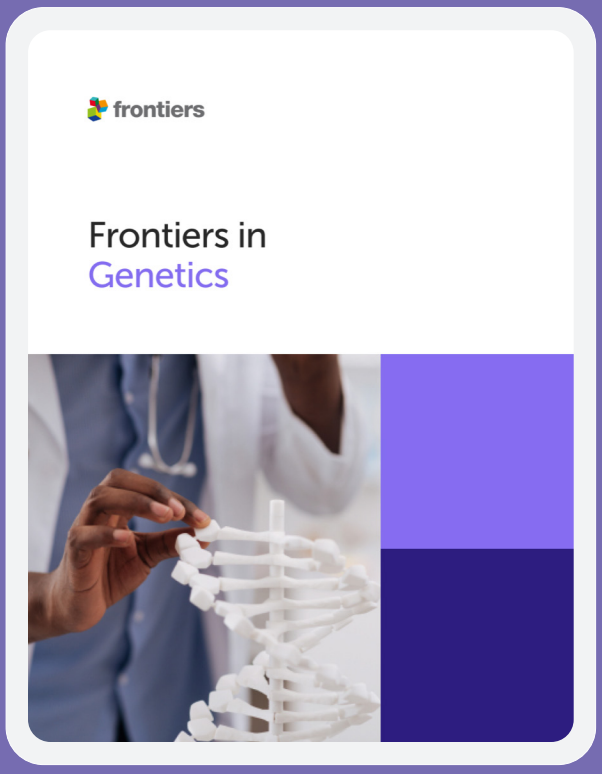
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact