Heliyon 9 (2023) e23029

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

# Genomic epidemiology reveals early transmission of SARS-CoV-2 and mutational dynamics in Nanning, China

DeWu Bi [a,e,f,*], XiaoLu Luo [a,e,f], ZhenCheng Chen [g], ZhouHua Xie [b,f], Ning Zang [h], LiDa Mo [a,f], ZeDuan Liu [a,f], YanRong Lin [d,f], YaQin Qin [d,f], XiKe Tang [d,f], Lü Lin [c,f], YuanLi Wang [g], LiangLi Cao [g], FeiJun Zhao [g], JinAi Zhou [h], ShanQiu Wei [a,f], ShaoYong Xi [a,f], QiuYing Ma [a,f], JianYan Lin [e,f,**]

[a] Department of Clinical Laboratory, The Fourth People's Hospital of Nanning, Nanning, China
[b] Department of Respiratory Medicine, The Fourth People's Hospital of Nanning, Nanning, China
[c] Emergency Department, The Fourth People's Hospital of Nanning, Nanning, China
[d] Department of Critical Care Medicine, The Fourth People's Hospital of Nanning, Nanning, China
[e] Key Laboratory of Infectious Diseases, The Fourth People's Hospital of Nanning, Nanning, China
[f] Affiliated Infectious Disease Hospital of Nanning, Guangxi Medical University, Nanning, China
[g] School of Life and Environmental Sciences, Guilin University of Electronic Technology, Guilin, China
[h] Guangxi Medical Research Center, Guangxi Medical University, Nanning, China

ARTICLE INFO

*Keywords:*
SARS-CoV-2
COVID-19
Variant
Evolution
Mutation

ABSTRACT

*Background:* Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants are a fatal pathogen resulting in substantial morbidity and mortality, and posing a great threat to human health with epidemics and pandemics. Methods: Next-generation sequencing (NGS) was performed to investigate the SARS-CoV-2 genomic characterization. Phylogenetic analysis of SARS-CoV-2 genomes was used to probe the evolutionary. Homology protein structure modelling was done to explore potential effect of the mutations. Results: The eighty genome sequences of SARS-CoV-2 obtained from the thirty-nine patients with COVID-19. A novel variant with mutation H625R concomitant with S50L in spike glycoprotein had been identified. Phylogenetic analysis revealed that SARS-CoV-2 variants belong to several distinct lineages. Homology modelling indicated that variant with mutation H625R and S50L increases flexibility of S1 subunit. Conclusions: SARS-CoV-2 genomes are constantly evolving by accumulation of point mutations. The amino acid H625R in combination with S50L may have a significant impact on the interaction between spike glycoprotein and ACE2.

## 1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the cause of the coronavirus disease 2019 (COVID-19) pandemic [1–3], is a highly contagious disease with high morbidity and mortality [3], which caused more than 259 million conformed cases and

* Corresponding author. The Fourth People's Hospital of Nanning, No.1, Erli, Changgang Road, Xingning District, 530023, Nanning, Guangxi, China.
** Corresponding author. Key Laboratory of Infectious Diseases, The Fourth People's Hospital of Nanning, Nanning, China.
*E-mail addresses:* dewubi@163.com (D. Bi), linjianyan@126.com (J. Lin).

https://doi.org/10.1016/j.heliyon.2023.e23029
Received 4 August 2023; Received in revised form 16 November 2023; Accepted 24 November 2023
Available online 29 November 2023
2405-8440/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

resulted in more than five million fatalities [4], since COVID-19 was first reported in China [2,5,6]. In line with previous studies on the causative mechanism suggested that human angiotensin converting enzyme 2 (hACE2), as a attachment receptor for entry into host cells [7–9], is responsible for the receptor binding domain (RBD)-mediated SARS-CoV-2 fusion into host cell membrane [7–11].

Strikingly, the ongoing evolution and emergence of SARS-CoV-2 variants presents a large number of challenges to global health [12–16], and may appear to decrease or invalidate the current monoclonal antibodies therapies and vaccines [13,17–24], which has attracted considerable concerns worldwide. Several lines of evidence suggest that the emergence of SARS-CoV-2 variants with mutations in the SARS-CoV-2 spike (S) glycoprotein, especially in RBD, have an adverse effect on humoral immunity induced by infection or vaccination [18–25].

An increasing number of studies have focused on the evolutionary consequences of spike protein with different point mutations [14,15,25–27], and the neutralization capacity of antibodies against SARS-CoV-2 [20,22–24], however, SARS-CoV-2 has been constantly evolving by accumulating novel mutations to evade immune surveillance. Therefore, additional studies are needed to better understand and address these issues.

In the present study, we identified a SARS-CoV-2 variant with H625R mutation, rarely reported in early 2020, concomitant with S50L mutation in spike glycoprotein. These findings implicate that targeting H625R and S50L mutation may serve as a potential allosteric site for variants susceptibility and guide vaccine design.

## 2. Materials and methods

### 2.1. Viral genome sequencing

The whole genome sequence of SARS-CoV-2 was amplified using the Ion AmpliSeq™ DNA custom Panel WG00428_Coronavirus (ThermoFisher, Waltham, MA, USA), containing two pools with 121 pair of primers. Viral cDNA was synthesized with the RNA-seq cDNA Synthesis Kit (ThermoFisher, Waltham, MA, USA), and then cDNA libraries were constructed with use of the Ion AmpliSeq Library Kit 2.0 (ThermoFisher, Waltham, MA, USA). Adaptor ligations were performed by use of the Ion Xpress Barcode Adapters Kit (ThermoFisher, Waltham, MA, USA), then purified with Agencourt AMPure XP beads (Beckman Coulter, Brea, CA, USA), followed by quantification with an Invitrogen Qubit 3.0 Fluorometer. The resulting DNA libraries were sequenced on Ion Proton Instrument (ThermoFisher, Waltham, MA, USA) using the Ion PI Hi-Q Sequencing 200 Kit (ThermoFisher, Waltham, MA, USA), then executed the whole genome assembly with the Iterative Refinement Meta-Assembler (IRMA) software (version 1.0) [28]. For this investigation, we only considered non-synonymous mutations. The deposited sequencing data generated from this study can be accessed on GISAID with the following accession ID: EPI_ISL_6951105, EPI_ISL_6951027, EPI_ISL_6951096, EPI_ISL_6951100, EPI_ISL_6951092, EPI_ISL_6951129, EPI_ISL_6951134, EPI_ISL_6951133, EPI_ISL_6951122, EPI_ISL_6951248, EPI_ISL_6951249, EPI_ISL_6951252, EPI_ISL_6951243, and EPI_ISL_6951257.

### 2.2. Sample collection

A total of 309 clinical specimens, including oropharynx swab, stool samples, sputum, and urine samples, were collected from 39 patients with COVID-19. Each of the specimens' collection was separated by at least 24 h. All cases were admitted, within the period from January 1st to March 30th, 2020, to the Fourth Peoples' Hospital of Nanning.

### 2.3. Homology modelling

Fourteen severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolates collected from February 1st, 2020, to March 30th, 2020, in Nanning reveals a novel variant that share at least 2 of 4 mutations (S: S50L; H625R; D614G; C1250F), designated as NAN. 20. Homology modelling of the SARS-CoV-2 wild type was performed via using Swiss-Model Expasy (www.swissmodel.expasy.org) and NAN. 20 variant on the basis of the Cry-EM structure of triple ACE2-bound SARS-CoV-2 trimer spike (PDB ID: 7KMS) [29], obtained from Protein Data Bank (https://www.rcsb.org), and served as a template derived from chain A. Structures were imported into UCSF Chimera v 1.15 after the homology modelling and single amino acid point mutations were inserted into the spike glycoprotein structure using the swapaa command. These mutation sites were S50L and H625R.

### 2.4. Phylogenetic analysis

Alignment of nucleotide and protein sequences were performed by Vector NTI Advance 11.0 (Invitrogen, USA). The phylogenetic tree was constructed by the Nextclade (sequence analysis webapp) in Nextstrain (https://nextstrain.org/sars-cov-2/).

### 2.5. Statistical analysis

The proportions of categorical variables were compared using the $\chi 2$ test. Statistical analyses were conducted using GraphPad Prism software (version 9·3, Graph-Pad Software, San Diego, California USA, https://www.graphpad.com/). A $p$ value less than 0.05 ($<0.05$) was considered significant.

### 2.6. Institutional review board statement

This was a retrospective study, and no participants were engaged in the design, setting, research objectives and questions, or clinical outcomes. There was no request for patients to provide interpretation or write up of results. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of The Ethics Committee of Fourth People's Hospital of Nanning (No. [2020]2, [2020]3, [2020]4, [2020]05).

## 3. Results

### 3.1. Baseline demographic and clinical characteristics

Of 39 patients with COVID-19 who were investigated, 54 % (21/39) were male. The diagnosis was made at a median age of 44 years (interquartile range, 35–65 years). Six hospitalized individuals have been diagnosed with severe COVID-19 illness symptoms and others had a diagnosis of moderate COVID-19. The first imported case who returned from Europe was reported on February 16, 2020.3.2.

### 3.2. Genomic characteristics of the NAN. 20 and related mutants

Genomic features of SARS-CoV-2, the seventh novel coronavirus known to infect human, shares with other coronavirus that the six functional open reading frames are presented with an order from 5′ end to 3' end: replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N) (Fig. 1). Alignment of the whole genome of the mutant NAN. 20 with the reference Wuhan-Hu-1 strain suggested that multiple non-synonymous were found in most open reading frames (Table 1). The conjunction analysis of the discovery and screening variants of SARS-CoV-2 confirmed the spike gene was the most frequently mutated gene, while ORF1b gene had lowest frequency of the mutations. According to the analysis, the mutation sites were predominantly located at the D614G and N501Y. Interestingly, these sites and adjacent more frequent mutation sites were mostly located in the receptor binding domain (RBD). Beyond this fact, nucleotide diversity analysis revealed that N gene had the highest single-site, G204R, frequency of the mutations.

### 3.3. Phylogenetic analysis

The Nextstrain phylogenetic tree analysis for the whole-genome of NAN. 20 variant and other mutations, revealed that the variant NAN.20/14-1, NAN.20/15, NAN.20/303, NAN.20/307, NAN.20/311, NAN.20/313-1, NAN.20/314, NAN.20/4, and NAN.20/13 belonged to the 19A clade; NAN.20/404 and NAN.20/408 related to 19B clade; NAN.20/214 arose from 20A clade, NAN.20/8 and NAN.20/1008 originated from 20B clade according to Nexstrain (Fig. 2). On the basic of the phylogenetic tree, NAN.20/8 and NAN.20/1008, as a distinct clade, forming an independent branch, were different from other NAN.20 variants. The sequences of NAN.20/8 and NAN.20/1008 variants were obtained from a patient who had been infected abroad ultimately return home from Europe during the COVID-19 pandemic, this data was consistent with epidemiological characteristics. Other NAN.20 variants were linked to the isolates in the early stage of the COVID-19 pandemic, which were associated with an earlier outbreak in Wuhan. These lines of evidence suggest that NAN. 20, except for the imported variants NAN.20, were the most prevalent variants of SARS-CoV-2 in Nanning, China.

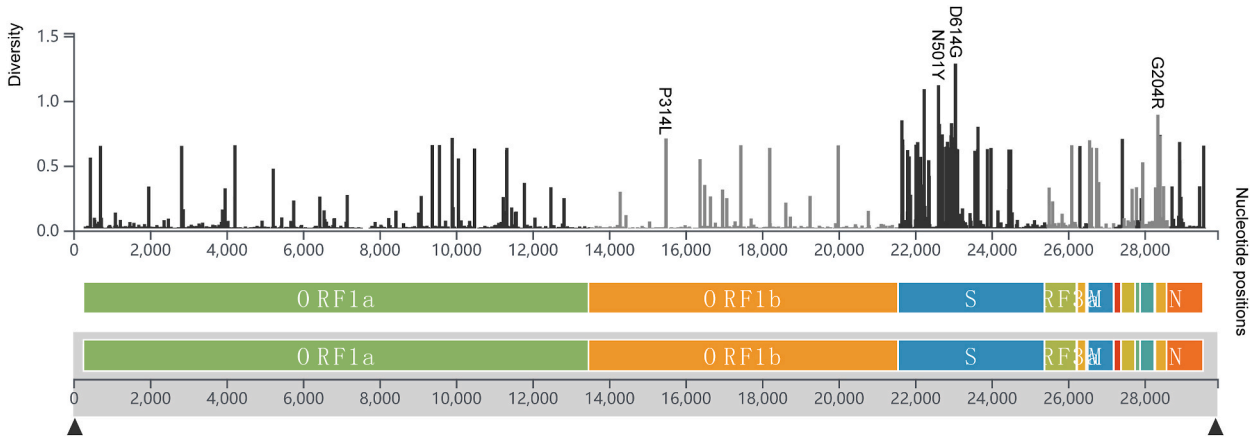


**Fig. 1.** nucleotide diversity observed in the variant of SARS-CoV-2. Amongst the single locus mutation, G204R mutation in N genes were found at the highest frequency, however, the gene with the highest mutation rate was spike, the high-frequency mutation of spike may be closely related to facilitating viral escape mechanisms.

**Table 1**
Lineage-defining protein altering mutations defining the NAN.20 SARS-CoV-2 variant.

| Gene | Nucleotide | Amino acid |
|---|---|---|
| ORF 1a | | |
| 1 | C635T | R124C |
| 2 | C2037T | A591V |
| 3 | T2330C | C689R |
| 4 | G2351A | G696S |
| 5 | C3885A | P1207H |
| 6 | C6312A | T2016K |
| 7 | A6315G | K2017R |
| 8 | C7164T | T2300I |
| 9 | T7872C | L2536S |
| 10 | T8289C | L2675P |
| 11 | G8561A | V2766I |
| 12 | A10380G | G3372V |
| 13 | A10874G | N3537D |
| 14 | G11083T | L3606F |
| 15 | G11930A | A3889T |
| 16 | G12131T | A3956S |
| 17 | G13127A | G4288R |
| ORF 1b | | |
| 1 | C13730T | A88V |
| 2 | C13945T | P160S |
| 3 | C14408T | P314L |
| 4 | A14702G | D412G |
| 5 | G16636A | A1057T |
| 6 | A18250G | N1595D |
| 7 | G21073A | V2536I |
| 8 | A21487G | R2674E |
| S | | |
| 1 | A23403G | D614G |
| 2 | C21711T | S50L |
| 3 | A23436G | H625R |
| 4 | G25311T | C1250F |
| ORF 3a | G26144T | G251V |
| ORF 8 | T28144C | L84S |
| ORF 9b | C28311T | P10S |
| N | | |
| 1 | R28882A | R203K |
| 2 | G28883C | G204R |
| 3 | T29445T | T391I |
| 4 | C28311T | P13L |
| 5 | G29276T | G335C |
| 6 | C29180T | Q303del |
| M | A26530G | D3G |

### *3.4. Mutation sites distribution of spike glycoprotein*

Notable characteristics of the spike glycoprotein is a polybasic cleavage site at the junction of subunits S1, responsible for receptor binding, and S2, liable for cell membrane fusion (Fig. 3a). Both subunits S1 and S2 harbour de novo mutations, in this study, except for D614G which was previously reported, three other non-synonymous spike glycoprotein mutations, namely S50L, H625R and C1250F, had been emerged. Three of the spike mutations, S50L, D614G and H625R, are in the S1 subunit, one, C1250F, is in the S2 subunit (Fig. 3b).

Overall, Chi-square contingency table analysis suggested that there was a significant difference in mutations between subunit S1 and S2 ($p < 0.001$). Mutation's occurrences are predominantly observed in subunits S1. Viral amino acid sequence analyses revealed a previously reported D614G mutant that is shared by all other emergence of novel SARS-CoV-2 variants. A significantly higher mutation rate was observed in another specific variant of SARS-CoV-2, S50L mutation, which collected shortly after a subsequent COVID-19 pandemic. The cumulative mutation rate of S50L was comparable with D614G, both S50L and D614G represent the largest epidemic strains in Nanning, China. Single site mutation occurrences were observed in D614G and C1250F, half of S50L were single-locus mutations, however, H625R was always concomitant with S50L.

### *3.5. Potential impact of spike glycoprotein variant NAN. 20*

To compare the interacting amino acid residues between the SARS-CoV-2 wild type and NAN. 20 variants, interested amino acid residues involved in the interaction, assigning contacts based on a distance criterion of less than or equal to 5 Å between any pair

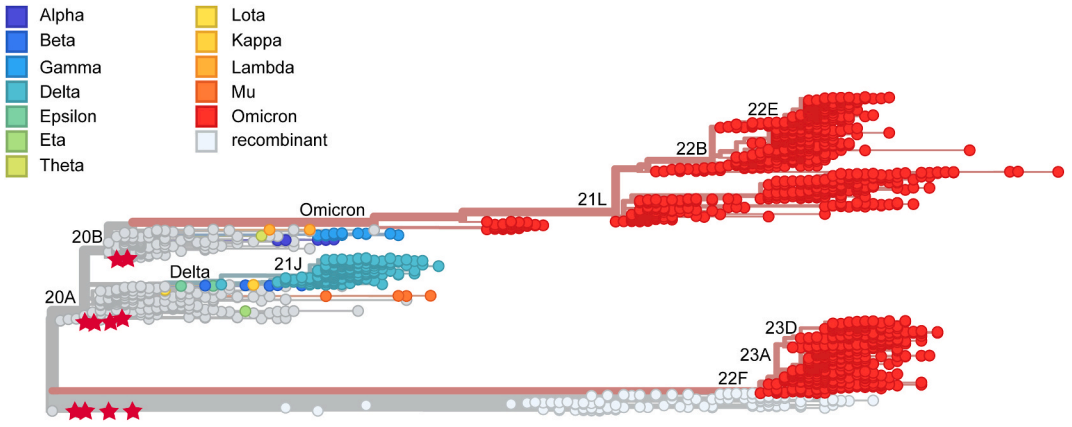


**Fig. 2.** Phylogenetic tree of the SARS-CoV-2 variants reconstruction based on Nextstrain tools. A phylogenetic tree of SARS-CoV-2 isolates, as well as a worldwide subsample of isolates collected between December 2019 and October 2021, are shown. The evolutionary tree depicts NAN. 20 links to other circulating lineages. The number of mutations acquired prior to discovery is shown in the branch length, and clades are labelled using Nextstrain nomenclature.

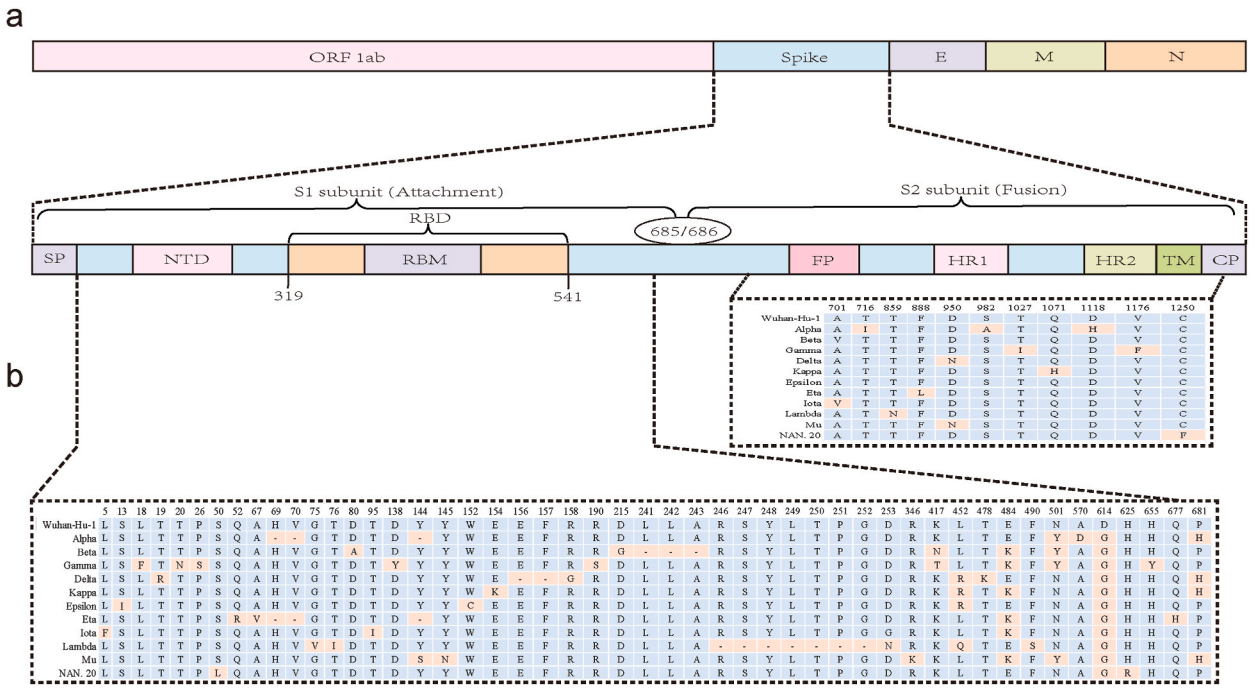


**Fig. 3.** Key differences in the spike glycoprotein of SARS-CoV-2 and the variants: (a) Schematic diagram of the spike glycoprotein of SARS-CoV-2. The residue numbers of each region correspond to their position in the spike glycoprotein of SARS-CoV-2. (b) Specific amino acid variations among the spike glycoproteins of Wuhan-Hu-1, Alpha, Beta, Gamma, Delta, Kappa, Epsilon, Eta, Iota, Lambda, Mu and NAN. 20. Characteristics of the important mutation residue in spike glycoprotein of novel SARS-CoV-2 variant was annotated (marked in orange). CP, cytoplasmic domain; FP, fusion peptide; HR1, heptad repeat 1; HR2, heptad repeat 2; NTD, N-terminal domain; SP, signal peptide; TM, transmembrane domain; RBD, receptor-binding domain; RBM, receptor-binding motif.

atoms, were depicted as sticks and indicated (Fig. 4a, b, c, d, e, f). In the N-terminal domain (NTD), mutated at 50 position leading to an amino acid substitution of serine (polar) to leucine (non-polar) induced a reduced flexibility for amino acid residues cysteine 291 (Cys291), cysteine 301 (Cys301) and threonine 302 (Thr302) (Fig. 4b, e). These residues are all located at the medial surface of the trimer and are uncharged polar amino acid residues. Compared to the wild type spike glycoprotein, the polar serine residue was substituted with non-polar leucine in the mutation spike glycoprotein induces an increase of conformational stability that the hydrophobicity of leucine has a lower free energy. Additionally, interactions of leucine 50 with surrounding residues cysteine 291, cysteine 301 and threonine 302, arguing for a stable conformation of this attachment peptide adjacent region, as it is in favour of by

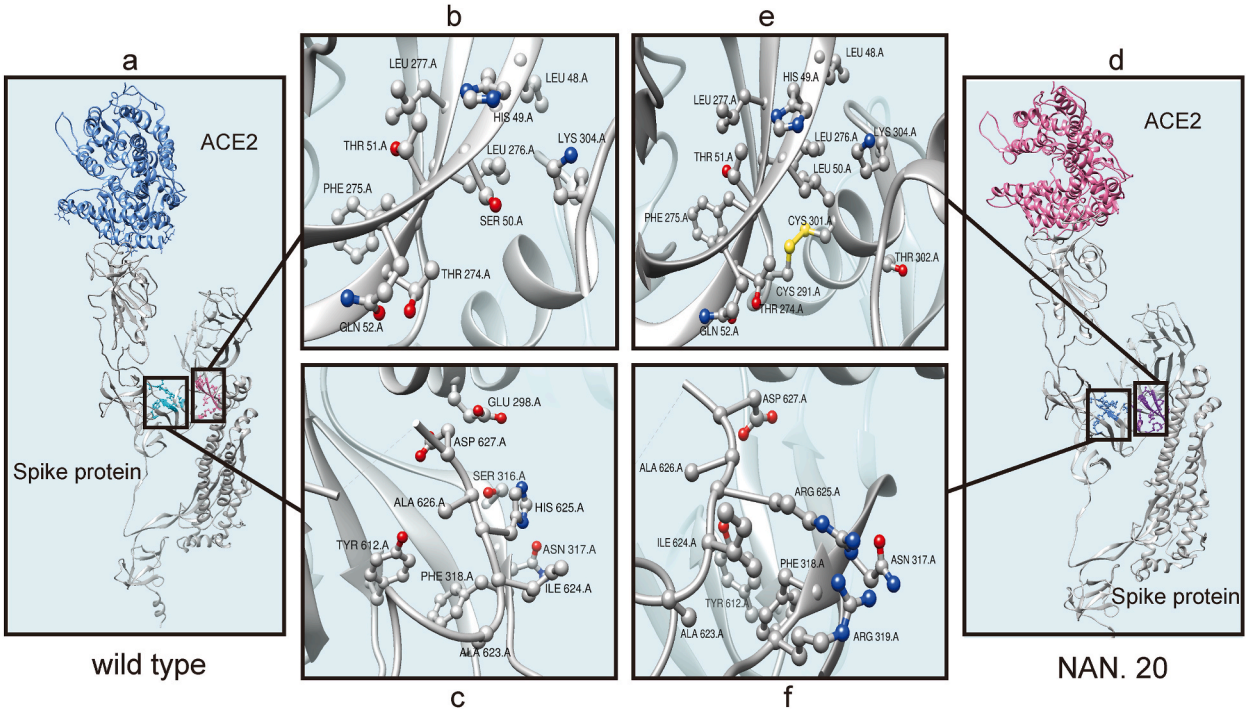


**Fig. 4.** Atomic details of mutation S50L and H625R interaction: (a) Structural representation of the spike glycoprotein (wild type) in complex with ACE2. (b) Interaction of the wild type spike glycoprotein amino acid residue (Ser 50) with surrounding residues leucine 48 (Leu48), histidine 49 (His49), threonine 51 (Thr51), glutamine 52 (Gln52), threonine 274 (Thr274), phenylalanine 275 (Phe275), leucine 276 (Leu276), leucine 277 (Leu277) and lysine 304 (Lys304). (c) Interaction of the wild type of spike glycoprotein amino acid residue (His 625) with surrounding residues glutamicacid 298 (Glu 298), serine 316 (Ser 316), asparagine 317 (Asn317), phenylalanine 318 (Phe318), tyrosine 612 (Tyr612), alanine 623 (Ala623), isoleucine 624 (Ile624), alanine 626 (Ala626) and asparticacid 627 (Asp627). (d) Structural representation of the spike glycoprotein (mutation NAN. 20) in complex with ACE2. (e) Interaction of the wild type of spike glycoprotein amino acid residue (Leu 50) with surrounding residues leucine 48 (Leu48), histidine 49 (His49), threonine 51 (Thr51), glutamine 52 (Gln52), threonine 274 (Thr274), phenylalanine 275 (Phe275), leucine 276 (Leu276), leucine 277 (Leu277) cysteine 291 (Cys291), cysteine 301 (Cys301), threonine 302 (Thr302) and lysine 304 (Lys304). (f) Interaction of the wild type of spike glycoprotein amino acid residue (Arg 625) with surrounding residues asparagine 317 (Asn317), phenylalanine 318 (Phe318), arginine 319 (Arg 319), tyrosine 612 (Tyr612), alanine 623 (Ala623), isoleucine 624 (Ile624), alanine 626 (Ala626) and asparticacid 627 (Asp627).

more three residues interaction than wild type spike glycoprotein.

The structural modelling investigation clearly showed that the amino acid 625 and surrounding residues, which interacted closely, were predominantly positioned at the surface of the spike glycoprotein, and a significant proportion of residues were discovered to interact with other spike glycoproteins at the viral surface (Figs. S1c and d). Interactions between histidine 625 and glutamicacid 298, histidine 625 and serine 316 were not identified in the variation NAN. 20 However, as compared to the wild type, mutant arginine 625 was closely connected with arginine 319, where the residue was positioned at the interface between two neighbouring spike trimers (Fig. 4c, f). To try to investigate differences between wild type and NAN. 20, structure analysis suggested that both wild type and variant NAN.20 residues asparagine 317, phenylalanine 592 at chain A interacted with asparticacid 737, threonine 739 and methionine 740 at chain B in the preferred distance of under 5 Å. (Figs. S1c and d). Regarding the NAN.20 variant, discovery of the residues arginine 319 at chain A contributed to interaction with asparagine 745 at chain B in a preferred distance of 4.734 Å when compared to wild type spike glycoprotein (Fig. S1d). Consequently, the mutant spike glycoprotein induced an increase of conformation stability.

To further probe the relationship between residue serine 50 and histidine 625, including pre- and post-mutated, we also analyzed interactions of this residues with surrounding amino acid. Interestingly, the residue serine 50 interacted with histidine 625 via seven amino acid residues, while leucine 50 associated with arginine 625 by four amino acid residues in the preferred distance of under 5 Å (Figs. S2a and b). Taken together, a decreased number of contacts for mutated leucine at site 50 and arginine at position 625 induced a loss of conformation stability that increased flexibility of this attachment peptide, which also explained why the mutation leucine 50 was concomitant with mutation arginine 625.

## 4. Discussion

In this study, a number of mutated sites had been identified in the SARS-CoV-2 diverse genomic locations. This result is consistent with previous studies showing genetic diversity by mutations [14,15,25,26], we found that gene S was mutated most frequently in all

open reading frames (ORFs) and the most commonly mutation position was D614G in gene S, while gene N captured the most commonly mutated site G204R in all ORFs. However, a novel mutation site, H625R in gene S, had been identified, which has rarely reported in databases or studies in early 2020. Furthermore, particularly intriguing is evidence that H625R appear to prefer concomitant with position mutation S50L, which suggests that they may improve attachment function between spike glycoprotein and receptor ACE2 by increasing flexibility in S1 subunit.

Thus, extensive investigations of spike glycoprotein had identified a strong preference for mutation in S1 subunit, with some preference in S2 subunit. Deletion and point mutation analyses suggested that the sites on spike glycoprotein vary but are primarily located in regions, which is outside of the RBD. The viral genetic diversity varied substantially among ORFs, with gene S maintaining particularly high level of genetic diversity. The earliest emergence of mutation site D614G, E484K mutations comprise the most frequently mutated hotspot in the receptor binding motif (RBM), followed by N501Y and L452R mutations [25,26,30], these mutations contribute not only to increasing transmissibility but also to decreasing the vaccine-induced neutralization. A large number of mutated sites were identified; however, no mutations were observed in RBD of NAN.20.

Phylogenetic analysis for the whole genome shows that variants NAN.20, divided into four different lineages, notably, on the Nextstrain-based phylogenies demonstrated that the SARS-CoV-2 sequence was frequently mutated, and acquired distinct patterns evolution, including linear and branching of evolution. Consistent with previous findings of evolution about mutational effect in evolutionary process, the SARS-CoV-2 has been continually evolving [14,25,26,30].

Recently, crystal structures of the complex of spike glycoprotein and receptor ACE2 have been solved [7–10]. On the basic of these prior studies, structure analysis reveal that mutation H625R is located on the surface of C-terminal domain (CTD) and S50L located on the surface of N-terminal domain (NTD). The CTD interacts with NTD by just a few amino acids, consistently, mutations H625R and S50L were found to be closely associated with these amino acids. Previous research found that nonsynonymous mutations, such as S50L in the S-protein, enhanced interaction between the S-protein and the ACE2 receptor, which may be linked to pathogenic high levels and long-term infection [31,32]. Thus, we speculate that mutations H625R and S50L might play an important role in SARS-CoV-2 susceptibility and represent a potential therapeutic target.

## 5. Conclusion

In conclusion, we illustrate two mutations in spike glycoprotein, simultaneously mutating sites S50L and H625R, and discuss the probable consequences of the variation NAN.20.

## Funding

This research was funded by The Scientific Research and Technology Development Program of Guangxi (grant number 2020AB39270 and 2020AB39022) and Guangxi Zhuang Region Health Department Funds (grant number Z2020856 and GZZC2020454).

## Data availability

Data associated with our study had been deposited into a publicly available repository. All the sequenced genomes were submitted to the GISAID database (https://gisaid.org/). Deposited sequences are available with accession number ID: EPI_ISL_6951105, EPI_ISL_6951027, EPI_ISL_6951096, EPI_ISL_6951100, EPI_ISL_6951092, EPI_ISL_6951129, EPI_ISL_6951134, EPI_ISL_6951133, EPI_ISL_6951122, EPI_ISL_6951248, EPI_ISL_6951249, EPI_ISL_6951252, EPI_ISL_6951243, and EPI_ISL_6951257.

## CRediT authorship contribution statement

**DeWu Bi:** Writing - original draft, Investigation, Conceptualization. **XiaoLu Luo:** Resources. **ZhenCheng Chen:** Funding acquisition. **ZhouHua Xie:** Funding acquisition. **Ning Zang:** Visualization, Software. **LiDa Mo:** Funding acquisition. **ZeDuan Liu:** Funding acquisition. **YanRong Lin:** Supervision, Investigation, Data curation. **YaQin Qin:** Validation, Methodology, Data curation. **XiKe Tang:** Visualization, Investigation. **Lü Lin:** Validation, Investigation, Data curation. **YuanLi Wang:** Software, Resources. **LiangLi Cao:** Software, Resources. **FeiJun Zhao:** Software, Resources. **JinAi Zhou:** Software, Resources. **ShanQiu Wei:** Supervision, Project administration. **ShaoYong Xi:** Supervision, Project administration. **QiuYing Ma:** Supervision, Project administration. **JianYan Lin:** Validation, Project administration, Conceptualization.

## Declaration of competing interest

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

## Acknowledgments

We gratefully thank the authors listed in Table S1, who had shared their genomic sequences of SARS-CoV-2 in this research. Meanwhile, technical support from Guangzhou Darui Biotechnology Co, Ltd is gratefully acknowledged.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e23029.

## References


[1] J.F. Chan, S. Yuan, K. Kok, K.K. To, H. Chu, J. Yang, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, Lancet 395 (2020) 514–523, https://doi.org/10.1016/S0140-6736(20)30154-9.

[2] N. Zhu, D. Zhang, D. Wang, W. Wang, X. Li, B. Yang, et al., A novel coronavirus from patients with pneumonia in China, 2019, N. Engl. J. Med. 382 (2020) 727–733, https://doi.org/10.1056/NEJMoa2001017.

[3] B. Hu, H. Guo, P. Zhou, Z. Shi, Characteristics of SARS-CoV-2 and COVID-19, Nat. Rev. Microbiol. 19 (2021) 141–154, https://doi.org/10.1038/s41579-020-00459-7.

[4] Who, Weekly operational update on COVID-19, Situation reports, https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19—23-november-2021.

[5] F. Wu, S. Zhao, B. Yu, Y. Chen, W. Wang, Z. Song, et al., A new coronavirus associated with human respiratory disease in China, Nature 579 (2020) 265–269, https://doi.org/10.1038/s41586-020-2008-3.

[6] P. Zhou, X. Yang, X. Wang, B. Hu, L. Zhang, W. Zhang, et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature 579 (2020) 270–273, https://doi.org/10.1038/s41586-020-2012-7.

[7] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, et al., Structural and functional basis of SARS-CoV-2 entry by using human ACE2, Cell 181 (2020) 894–904, https://doi.org/10.1016/j.cell.2020.03.045.

[8] R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2, Science 367 (2020) 1444–1448, https://doi.org/10.1126/science.abb2762.

[9] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, et al., Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (2020) 221–224, https://doi.org/10.1038/s41586-020-2179-y.

[10] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (2020) 215–220, https://doi.org/10.1038/s41586-020-2180-5.

[11] A.C. Walls, Y. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, function, and Antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292, https://doi.org/10.1016/j.cell.2020.02.058.

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

[13] D. Planas, T. Bruel, L. Grzelak, F. Guivel-Benhassine, I. Staropoli, F. Porrot, et al., Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies, Nat Med 27 (2021) 917–924, https://doi.org/10.1038/s41591-021-01318-5.

[14] W. Zhang, B.D. Davis, S.S. Chen, J.M. Sincuir Martinez, J.T. Plummer, E. Vail, Emergence of a novel SARS-CoV-2 variant in Southern California, JAMA 325 (2021) 1324, https://doi.org/10.1001/jama.2021.1612.

[15] E. Callaway, Heavily mutated Omicron variant puts scientists on alert, Nature (2021), https://doi.org/10.1038/d41586-021-03552-w.

[16] H. Tegally, E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, et al., Detection of a SARS-CoV-2 variant of concern in South Africa, Nature 592 (2021) 438–443, https://doi.org/10.1038/s41586-021-03402-9.

[17] Y. Liu, J. Liu, H. Xia, X. Zhang, C.R. Fontes-Garfias, K.A. Swanson, et al., Neutralizing Activity of BNT162b2-elicited serum, N. Engl. J. Med. 384 (2021) 1466–1468, https://doi.org/10.1056/NEJMc2102017.

[18] C. Lucas, C.B.F. Vogels, I. Yildirim, J.E. Rothman, P. Lu, V. Monteiro, et al., Impact of circulating SARS-CoV-2 variants on mRNA vaccine-induced immunity, Nature (2021), https://doi.org/10.1038/s41586-021-04085-y.

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

[20] W.F. Garcia-Beltran, E.C. Lam, St Denis K, A.D. Nitido, Z.H. Garcia, B.M. Hauser, et al., Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity, Cell 184 (2021) 2372–2383, https://doi.org/10.1016/j.cell.2021.03.013.

[21] Z. Wang, F. Schmidt, Y. Weisblum, F. Muecksch, C.O. Barnes, S. Finkin, et al., mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants, Nature 592 (2021) 616–622, https://doi.org/10.1038/s41586-021-03324-6.

[22] R.E. Chen, X. Zhang, J.B. Case, E.S. Winkler, Y. Liu, L.A. VanBlargan, et al., Resistance of SARS-CoV-2 variants to neutralization by monoclonal and serum-derived polyclonal antibodies, Nat Med 27 (2021) 717–726, https://doi.org/10.1038/s41591-021-01294-w.

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

[24] L. Stamatatos, J. Czartoski, Y. Wan, L.J. Homad, V. Rubin, H. Glantz, et al., mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection, Science 372 (2021) 1413–1418, https://doi.org/10.1126/science.abg9175.

[25] H.S. Vöhringer, T. Sanderson, M. Sinnott, N. De Maio, T. Nguyen, R. Goater, et al., Genomic reconstruction of the SARS-CoV-2 epidemic in England, Nature (2021), https://doi.org/10.1038/s41586-021-04069-y.

[26] K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of SARS-CoV-2, Nat Med 26 (2020) 450–452, https://doi.org/10.1038/s41591-020-0820-9.

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

[28] S.S. Shepard, S. Meno, J. Bahl, M.M. Wilson, J. Barnes, E. Neuhaus, Viral deep sequencing needs an adaptive approach: IRMA, the iterative refinement meta-assembler, BMC Genom. 17 (2016) 1–18, https://doi.org/10.1186/s12864-016-3030-6.

[29] T. Zhou, Y. Tsybovsky, J. Gorman, M. Rapp, G. Cerutti, G. Chuang, et al., Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-Dependent Switch to mediate Endosomal positioning of receptor-binding domains, Cell Host Microbe 28 (2020) 867–879, https://doi.org/10.1016/j.chom.2020.11.004.

[30] D. Desai, A.R. Khan, M. Soneja, A. Mittal, S. Naik, P. Kodan, et al., Effectiveness of an Inactivated Virus-Based SARS-CoV-2 Vaccine, BBV152, in India: a Test-Negative, Case-Control Study, The Lancet Infectious Diseases, 2021, https://doi.org/10.1016/S1473-3099(21)00674-5.

[31] R.Y. Aljindan, A.M. Al-Subaie, A.I. Al-Ohali, et al., Investigation of nonsynonymous mutations in the Spike protein of SARS-CoV-2 and its interaction with the ACE2 receptor by molecular docking and MM/GBSA approach[J], Comput. Biol. Med. (2021), 104654, https://doi.org/10.1016/j.compbiomed.2021.104654.

[32] V. Borges, J. Isidro, M. Cunha, et al., Long-Term Evolution of SARS-CoV-2 in an Immunocompromised Patient with Non-hodgkin Lymphoma[J], mSphere, 2021, https://doi.org/10.1128/mSphere.00244-21.