



# A Proofreading Mutation with an Allosteric Effect Allows a Cluster of SARS-CoV-2 Viruses to Rapidly Evolve

Andrew H. Mack <sup>1</sup>, Georgina Menzies,<sup>1</sup> Alex Southgate,<sup>1</sup> D. Dafydd Jones,<sup>\*,1</sup> and Thomas R. Connor <sup>\*,1,2</sup>

<sup>1</sup>Molecular Biosciences Division, School of Biosciences, Cardiff University, UK

<sup>2</sup>Pathogen Genomics Unit, Public Health Wales NHS Trust, Cardiff, UK

\*Corresponding authors: E-mails: JonesDD@cardiff.ac.uk; ConnorTR@cardiff.ac.uk.

Associate editor: Thomas Leitner

## Abstract

**The RNA-dependent RNA polymerase of the severe acute respiratory syndrome coronavirus 2 virus is error prone, with errors being corrected by the exonuclease (NSP14) proofreading mechanism. However, the mutagenesis and subsequent evolutionary trajectory of the virus is mediated by the delicate interplay of replicase fidelity and environmental pressures. Here, we have shown that a single, distal mutation (F60S) in NSP14 can have a profound impact upon proofreading with an increased accumulation of mutations and elevated evolutionary rate being observed. Understanding the implications of these changes is crucial, as these underlying mutational processes may have important implications for understanding the population-wide evolution of the virus. This study underscores the urgent need for continued research into the replicative mechanisms of this virus to combat its continued impact on global health, through the re-emergence of immuno-evasive variants.**

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the coronavirus disease 2019 (COVID-19) global pandemic (Lu et al. 2020). SARS-CoV-2 is a single-strand positive sense RNA virus belonging to the family *Coronaviridae*, genera *Betacoronavirus*, to which the other notable pathogenic human coronaviruses (OC43, HKU1, SARS-CoV-1, and MERS-CoV) also belong (Lu et al. 2020). One important aspect of the behavior of all viruses is their capacity for genetic diversity through mutations. Mutations can arise from errors during the replication of the viral genome, and they have the potential to alter the properties of a virus in various ways. These changes can range from having no effect, through to altering the severity of the virus, rendering it resistant to antiviral drugs, or influencing its transmissibility.

RNA viruses, like SARS-CoV-2, have an estimated mutation rate of between  $10^{-5}$  and  $10^{-2}$  substitutions/per site/per year (s/s/y) (Duffy et al. 2008), which is significantly higher than that observed in DNA viruses which typically have mutation rates in the order of  $10^{-8}$ – $10^{-6}$  s/s/y (Sanjuán et al. 2010; Peck and Luring 2018). In comparison with organisms with DNA-based genomes, RNA viruses of the *Coronaviridae* family do not utilize high-accuracy polymerases. Instead, they utilize the RNA-dependent RNA polymerase (RdRp), which, on its own, operates at a faster pace but with a significantly higher error rate compared with most viruses (Shannon et al. 2020). Therefore, coronaviruses employ an independent proofreading mechanism

in the form of an exonuclease to correct errors introduced by the RdRp. The exonuclease of SARS-CoV-2 is housed within the nonstructural protein 14 (NSP14) coding segment, along with an N7-guanine methyltransferase. This exonuclease serves a crucial function in maintaining the accuracy of the viral RNA genome by removing erroneous nucleotides during replication. This process helps reduce the rapid accumulation of mutations which could be deleterious, ensuring that most virions generated are viable, and allows the virus to maintain its relatively large genome, which is ~30 kb in size, while still functioning efficiently (Gorbalenya et al. 2006; Ogando et al. 2020; Jo et al. 2021). Even with the proofreading capabilities of the exonuclease, SARS-CoV-2 has been able to adapt to the human host through the acquisition of mutations, exploring evolutionary space in a stepwise way. The rate of the emergence of variants/advantageous mutations is partly a function of the mutation rate of the virus, which is itself a function of the interaction of NSP14 and RdRp. It is also important to note that although mutations operate to generate new diversity in the population, coronaviruses are also able to undergo recombination, which serves to shuffle diversity within the wider population and has already been seen in SARS-CoV-2 (Jackson et al. 2021). This diversity can further confer an advantage which can then spread through lateral gene transfer into other contexts. Therefore, it is crucial to comprehend and dissect the processes responsible for mutations in SARS-CoV-2 in order to gain a deeper insight into the factors contributing to the emergence of diversity within the virus.

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Open Access

The exonuclease core of NSP14 within SARS-CoV-2 is a DEEDh enzyme and member of the DEED superfamily (Zuo and Deutscher 2001). The active site residues are comprised of two glutamic acid residues (E92 and E191), two aspartic acid residues (D90 and D273), and a single histidine residue (H268) which has been structurally confirmed and is similar to the NSP14 of SARS-CoV-1 (Ma et al., 2015; Liu et al. 2021; Lin et al. 2021). The excision of mis-incorporated nucleotides which are introduced through replication is facilitated by two coordinated  $Mg^{2+}$  ions (fig. 1). Like other exonucleases, the first  $Mg^{2+}$  ion (which is coordinated by E92 and D273) activates a water molecule to initiate nucleophilic attack, whereas the second  $Mg^{2+}$  ion (coordinated by D90 and E191) helps to remove the erroneous nucleotides from the replication complex (Hwang et al. 2018). Thus, coordination of  $Mg^{2+}$  ions play a critical role in maintaining the accuracy of the viral RNA genome (Ma et al., 2015).

Inactivation of the exonuclease enzyme of HCoV-229E and MERS-CoV through site-directed mutagenesis of key catalytic residues has previously been shown to prevent the successful generation of replication competent, viable virions (Minskaia et al. 2006; Ogando et al. 2020). However, although site-directed mutagenesis experiments of the DEEDh core of HCoV-229E and MERS-CoV rendered these viruses inviable, similar experiments conducted upon SARS-CoV-1 and murine hepatitis virus (MHV) produced viable virions, but with their fidelity greatly compromised, with progenitors displaying a significantly increased burden of mutations with an accompanying fitness cost (Eckerle et al. 2007, 2010; Graepel et al. 2017). Elsewhere, others have recognized that general mutations within the exonuclease are associated with an increased burden of mutations upon the SARS-CoV-2 viral genome (Eskier et al. 2020), but the underlying functional cause was not explored. In addition to its proofreading ability, the exonuclease has also been shown to be crucial for native recombination, as its inactivation resulted in a significant decrease in recombination frequency within MHV (Gribble et al. 2021), further highlighting the diverse roles of this enzyme.

In vitro reverse genetics experiments have shown that the SARS-CoV-1 exonuclease activity is mediated by its interaction with NSP10 (Bouvet et al. 2014). This work illustrated how conserved mutations of residues at the interface of the exonuclease and NSP10 within SARS-CoV-1 not only reduced exonuclease activity but were also capable of rendering the virus severely attenuated or nonviable. Enzymatic characterization experiments have recognized the significance of the NSP14/NSP10 interaction in modulating exonuclease activity, with the complex showing a >35-fold increase in activity compared with the NSP14 alone (Bouvet et al. 2012). This modulation of replication fidelity was further observed in vitro within a MHV model, where mutations at the NSP14/NSP10 interface led to an inability to excise the nucleotide analogues 5-azacytidine and ribavirin (Smith et al. 2015).

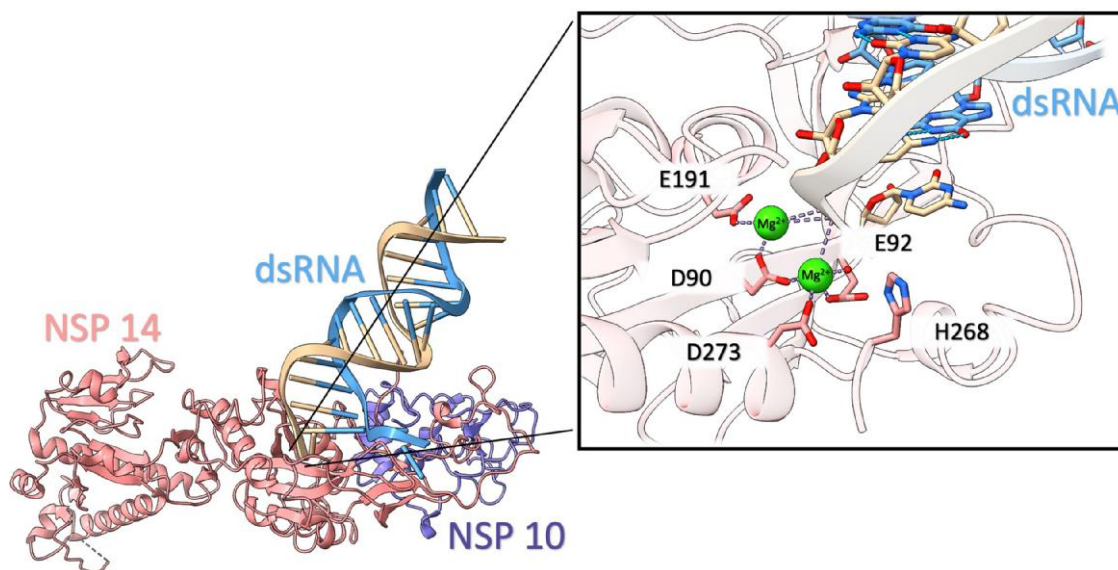
Considering the above, this investigation sought to explore whether mutations at the interface of the NSP14

and NSP10 resulted in demonstrable effects upon the evolutionary trajectory of the SARS-CoV-2 virus. Already, the global SARS-CoV-2 sequence data sets provide an unprecedented resource of genome sequence data that enables the identification of clusters of cases, variants of concern (VoC), and an estimation of their observed evolutionary rates. The large, global data set provides an opportunity to examine the generation of diversity in the SARS-CoV-2 population at unprecedented resolution. Building on the efforts of the global scientific community has allowed us to identify a cluster of SARS-CoV-2 cases that possess an interface mutation and to quantify that these viruses exhibit an elevated evolutionary rate. Using these mutations as a starting point, we undertook further in silico investigations to determine their potential impact on exonuclease function through predictive modeling, molecular dynamics (MD), and residue interaction network analyses, which have emerged as powerful tools in the investigation of the effect of mutations on replicase enzymes, including exonuclease function (Berger and Cisneros 2023). By modeling the structural dynamics of the exonuclease, we were able to determine whether mutations could lead to a distal allosteric effect on the catalytic residues of the enzyme, contributing to the observed higher mutation rate. The in-depth examination provided by these simulations offered crucial insights into the complex interaction between the mutations and the protein structure, corroborating our evolutionary findings and providing a more comprehensive understanding of the underlying mechanisms.

## Results

### Identifying Sequences with Mutations at the NSP14/NSP10 Interface

An examination of the NSP14/NSP10 interface through network analysis (as shown in [supplementary fig. 1](#), [Supplementary Material](#) online) uncovered interactions between 29 residues from NSP10 and 32 residues from NSP14. These interactions mainly consisted of Van der Waals forces (40 in total), with additional contributions from 8 hydrogen bonds, 4 Pi–Pi interactions, and 1 ionic bond. Prior research has noted two specific Pi–Pi interactions, between F16 of NSP10 and Y64 of NSP14 and F60 of NSP14 and F19 of NSP10, which have been considered as a possible therapeutic target due to their potential allosteric impact upon the catalytic core (Saramago et al. 2021) (fig. 2 shows the localization of NSP10 residues with NSP14). After identifying all the residues at the interface, a search was conducted on the Global Initiative on Sharing All Influenza Data (GISAID) (Shu and McCauley 2017) database to find SARS-CoV-2 sequences with mutations at those sites. As a result, 14 groups of SARS-CoV-2 sequences with mutations at the interface were discovered. Groups are subsequently named by the interface mutation they possessed. Additionally, two control lineages were obtained that did not possess any mutations at the NSP14/NSP10 interface. The complete list of data sets generated



**Fig. 1.** Structure of SARS-CoV-2 NSP14/NSP10 complex with dsRNA. This derives from the high-resolution crystal structure PDB:7N0C, with the E191A mutation reverted. The zoomed-in view of the catalytic core depicts the two  $Mg^{2+}$  ions that coordinate with the residues of the enzyme, providing insight into the key elements that facilitate the exonuclease activity of the NSP14. This figure provides a visual representation of the critical components of the complex and its interaction with dsRNA (PDB:7N0C).

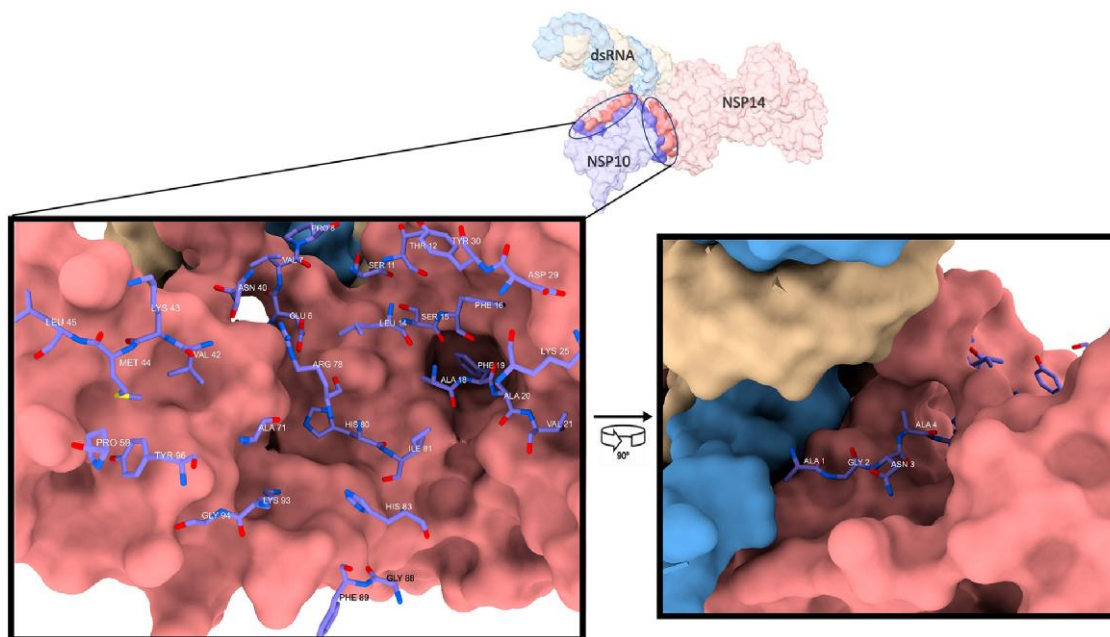
for this study, along with the final number of samples in each, is presented in [supplementary table S1, Supplementary Material](#) online.

#### Temporal Signal and Bayesian Evolutionary Analysis

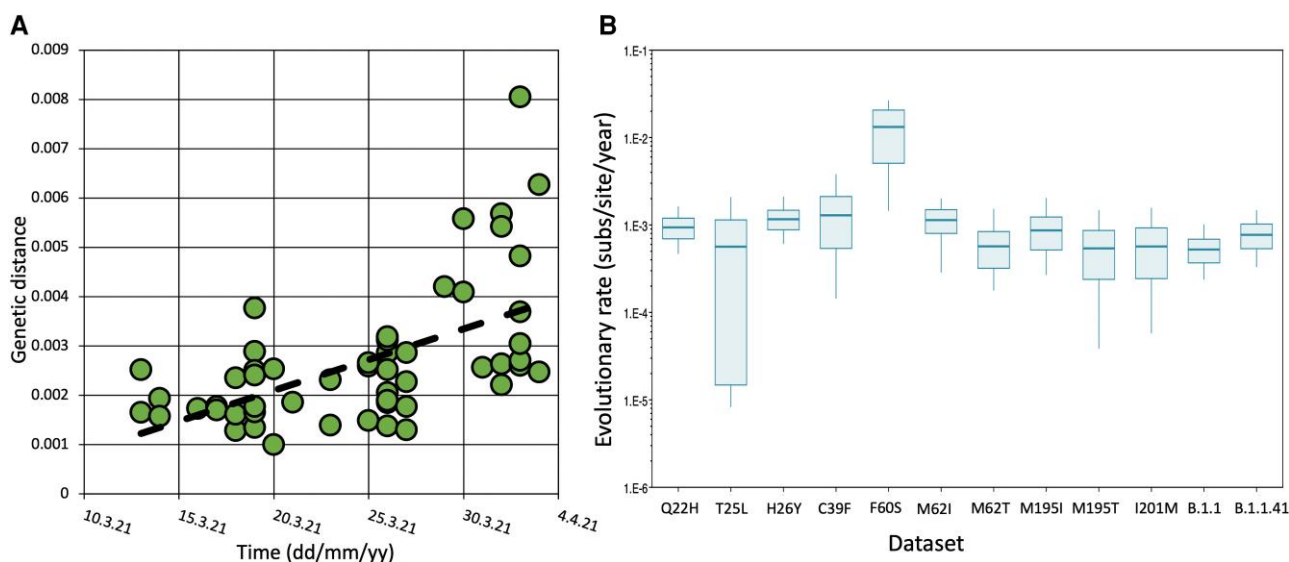
As an initial assessment of a temporal signal resembling a molecular clock within the identified control and mutant lineages, a linear root-to-tip regression was performed using TempEst v1.5.3 (Rambaut et al. 2016). This step was taken to verify that the data sets were suitable for further Bayesian analysis. Maximum likelihood phylogenies were created for each data set, highly divergent genomes were removed, and data set was rebuilt. A temporal signal was observed in all data sets tested, with rates ranging from  $1.86 \times 10^{-4}$  to  $2.63 \times 10^{-3}$  (supplementary table S2, Supplementary Material online). The  $R^2$  value, which indicates the degree to which a sample is “clocklike,” also showed variability. For example,  $R^2$  for the M62I mutant was 0.83, whereas F60S had a much lower  $R^2$  of 0.007, indicating a departure from clocklike behavior and a more variable rate of evolution. A root-to-tip regression for the F60S data set is shown in [figure 3A](#), whereas a complete data set of root-to-tip results is presented in [supplementary table S2, Supplementary Material](#) online. Having detected an initial temporal signal in all data sets, they were all subjected to further Bayesian analysis using BEAST v.1.10.4 (Suchard et al. 2018) to estimate their evolutionary rates. The most appropriate clock model for each data set was first determined through path sampling and stepping stone (SS) analysis (supplementary table S3, Supplementary Material online). All other priors were kept consistent across data sets, including the tree prior, as it was believed to have limited effect on the final molecular clock estimates (Ritchie et

al. 2017). Once the best model for each data set was identified, an additional and more robust test of temporal signal was conducted. A Bayesian evaluation of temporal signal (BETS) (Duchene et al. 2020) was performed on each data set, where the aforementioned BEAST analysis was repeated; however, sampling dates were removed. This analysis revealed that data sets H26N, M57I, M57V, and M62V possessed either a weak or no temporal signal, with a Bayes factor  $< 1$ , and consequently, any downstream inference of these data sets would not be reliable and were removed from the study (supplementary table S6, Supplementary Material online). Evolutionary rates of SARS-CoV-2 are commonly estimated to be within the range of  $7.0 \times 10^{-4}$  s/s/y (Ghafari et al. 2022). In this study, we found the evolutionary rate of our two comparator, nonmutant groups to also reside within or close to this estimate; we subsequently name these groups controls. Control group 1 (B.1.1 lineage) resulted in an s/s/y rate of  $5.27 \times 10^{-4}$  [95% highest posterior density (HPD)  $3.72 \times 10^{-4}$ – $6.92 \times 10^{-3}$ ], and Control group 2 (B.1.1.41) displayed an s/s/y rate of  $7.71 \times 10^{-4}$  (95% HPD  $5.35 \times 10^{-4}$ – $1.02 \times 10^{-3}$ ). In comparison, certain data sets containing mutations at the NSP14/NSP10 interface deviated strongly from this expected rate, particularly the F60S and C39F data sets, which resulted in s/s/y rates of  $2.37 \times 10^{-2}$  (95% HPD  $8.03 \times 10^{-3}$ – $3.83 \times 10^{-2}$ ) and  $4.06 \times 10^{-3}$  (95% HPD  $7.55 \times 10^{-4}$ – $8.57 \times 10^{-3}$ ), respectively. The posterior distribution of the evolutionary rates from the BEAST analyses is shown in [figure 3B](#). Furthermore, both data sets showed a coefficient of variation which had departed from and was no longer abutting zero, exemplifying how the evolutionary rate of both sets of genomes have departed from clocklike behavior, exhibiting variation between branch rates. We recognize that data





**FIG. 2.** NSP10 and NSP14 interaction in SARS-CoV-2—a representation of the localization of NSP10 residues to the surface model of NSP14 (PDB:7N0C).



**FIG. 3.** (A) Root-to-tip regression plot of the 55 F60S sequences showing genetic distance versus time (in dd/mm/yy). (B) Posterior distribution of the evolutionary rate ( $s/s/y$ ) of each data set analyzed under the best ranking model using BEAST.

sets containing fewer sequences tend to align with broader 95% HPD intervals, implying a reduced level of certainty in the estimates—as expected. Table 1 presents the results of the Bayesian analysis for all data sets.

#### Detection of Previous Recombination Event within the F60S Lineage

Although the C39F mutant exhibited a notably elevated evolutionary rate, the rate observed in the F60S data set was considerably higher and became the focus in

downstream analyses. To elucidate the possible cause for this dramatic elevation in the observed evolutionary rate, a test of recombination was performed. Using RDP5 (Martin et al. 2021), an initial analysis was conducted within the F60S data set, which did not yield any evidence of recombination. Following this, a second analysis was undertaken where 3,000 genomes from throughout the pandemic were obtained from GISAID. For computational efficiency, this data set was split into 10 tranches (~300 sequences per tranche) with the F60S data set being added to each tranche followed by another test of recombination

**Table 1.** Results of BEAST Analysis.

Data Set	Number of Sequences	Clock Model	Mean Rate (s/s/y)	95% HPD (s/s/y)
Q22H	100	E relaxed	$9.38 \times 10^{-4}$	$6.42 \times 10^{-4}$ – $1.25 \times 10^{-3}$
T25L	24	E relaxed	$5.98 \times 10^{-4}$	$4.17 \times 10^{-5}$ – $1.40 \times 10^{-3}$
H26Y	100	LN relaxed	$1.26 \times 10^{-3}$	$7.24 \times 10^{-4}$ – $1.91 \times 10^{-3}$
C39F	54	LN relaxed	$4.06 \times 10^{-3}$	$7.55 \times 10^{-4}$ – $8.57 \times 10^{-3}$
F60S	55	E relaxed	$1.36 \times 10^{-2}$	$8.51 \times 10^{-3}$ – $3.83 \times 10^{-2}$
M62I	102	LN relaxed	$1.58 \times 10^{-3}$	$7.11 \times 10^{-4}$ – $2.84 \times 10^{-3}$
M62T	100	E relaxed	$6.07 \times 10^{-4}$	$3.25 \times 10^{-4}$ – $9.16 \times 10^{-4}$
M195I	54	E relaxed	$1.01 \times 10^{-3}$	$5.36 \times 10^{-4}$ – $1.53 \times 10^{-3}$
M195T	55	E relaxed	$6.34 \times 10^{-4}$	$2.40 \times 10^{-4}$ – $1.08 \times 10^{-3}$
I201M	112	E relaxed	$6.28 \times 10^{-4}$	$2.91 \times 10^{-4}$ – $1.00 \times 10^{-3}$
Control 1 B.1.1	100	Strict	$5.27 \times 10^{-4}$	$3.72 \times 10^{-4}$ – $6.92 \times 10^{-3}$
Control 2 B.1.1.41	77	Strict	$7.71 \times 10^{-4}$	$5.35 \times 10^{-4}$ – $1.02 \times 10^{-3}$

using RDP5. From this analysis, results indicated that a recombination event was likely present within the F60S genomes, which was supported by 6/9 methods deployed in RDP5 ([supplementary table S4, Supplementary Material online](#)); these included MaxChi ( $P = <0.05$ ), 3Seq ( $P = <0.01$ ), RDP ( $P = <0.001$ ), GENECOV ( $P = <0.001$ ), and BootScan ( $P = <0.0001$ ). The recombination breakpoints were suggested to be at nucleotides 21,742 and 23,666 (codons 61–705 of the spike protein), with the inferred (unknown) major parent being from the B.1.1.529 lineage (Omicron) and minor parent being from the B.1.1 lineage, suggesting the B.1.1 minor parent had donated its S1 domain of the spike protein during this recombination event ([supplementary fig. 2, Supplementary Material online](#)). Although the F60S genomes were recombinant, this was not considered to be the cause of the elevated rate, as this event was basal to data set that was examined using BEAST, and therefore, it would have no bearing upon this analysis. However, this does highlight that these genomes, with a higher detectable rate of mutation, do have a history of recombination, pointing to the potential for these two mechanisms to combine to generate significant diversity within the population.

### MD Simulations of F60S and Wild-Type NSP14/NSP10 Complexes

Having established that there was no evidence pointing toward intra-dataset recombination as a cause of the elevated rate, we sought to specifically examine if the F60S mutation had any effect upon the NSP14/NSP10 protein complex of SARS-CoV-2. To examine whether F60S may influence its behavior, MD simulations were performed. The source structure for the MD was PDB:7N0C. This structure possessed the mutation E191A at the catalytic core for experimental purposes; this mutation was reverted to wild-type glutamate using CHARMM-GUI ([Jo et al. 2008](#)), which was also used to introduce the F60S mutation to the mutant starting structure. Simulations were ran for 100 ns with three repeats. Root mean squared deviation (RMSD) analysis suggested all simulations were stable ([supplementary fig. 3, Supplementary Material online](#)). An

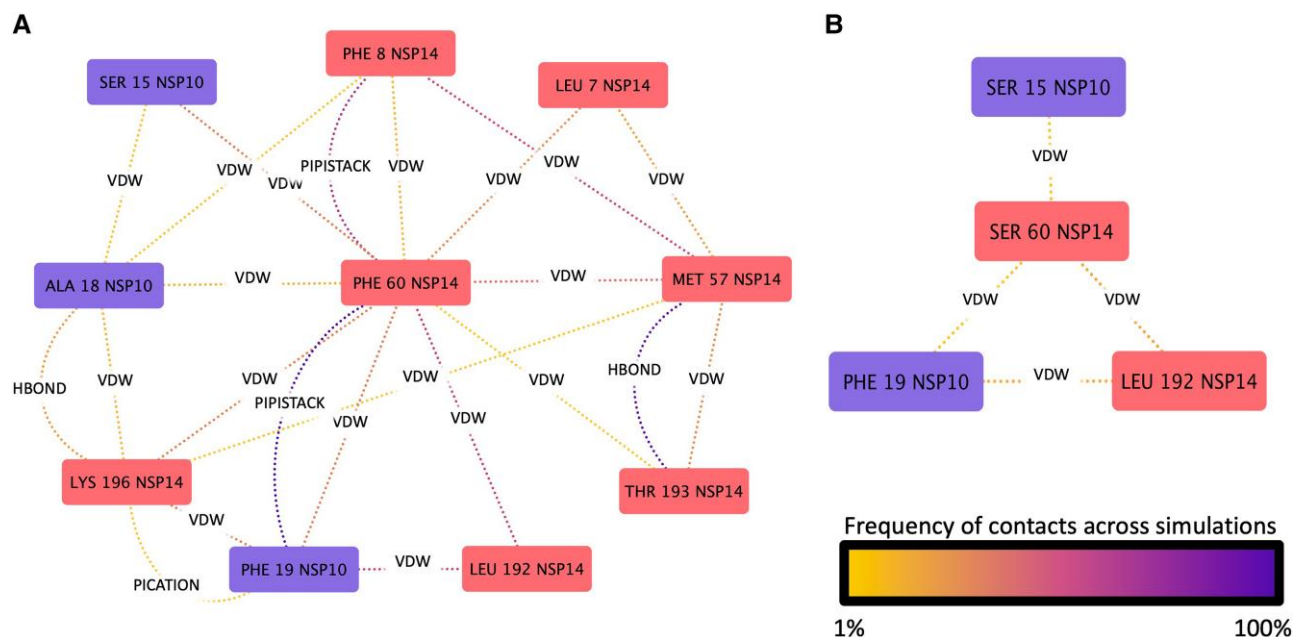
analysis of the F60S NSP14/NSP10 complex revealed the N-terminus of NSP10 that interacts with double-stranded RNA, and NSP10 was more flexible, but the stability between wild-type and mutant RNA was comparable. Within NSP14, although changes in root mean squared fluctuation (RMSF) were observed, particularly an increase in flexibility between residues 221 and 231, no significant changes were observed at the catalytic core ([supplementary fig. 4, Supplementary Material online](#)).

### Changes in the Residue Interaction Networks at the F60S Mutation Site

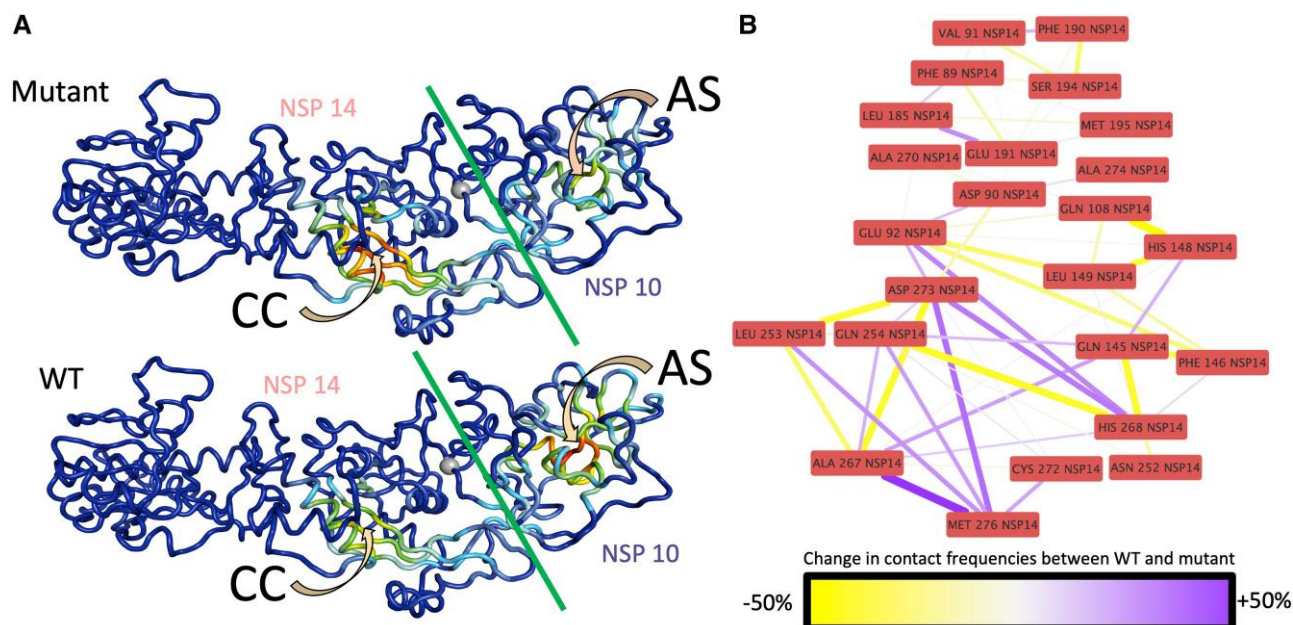
To better illustrate the changes in the local interactions within the proximity of the F60S mutation, the three separate 100 ns trajectories were concatenated to single 300 ns simulation for both the wild-type and mutant. Six hundred frames (20% of total) from each concatenated simulation were then extracted, and a residue interaction network was created using RING3.0 ([Clementel et al. 2022](#)). This network analysis clearly illustrates the local changes between the wild-type and mutant structures ([fig. 4](#)), where the F60S mutation led to the loss of seven out of the ten residue interactions seen in the wild-type and a dramatic decrease in the frequency of the remaining interactions.

### Predicting the Effect of F60S Mutation on the Binding Affinity of NSP14/NSP10 and Interface Surface Size

To determine the impact of the F60S mutation on the binding strength of NSP14 with NSP10, multiple structure-based tools including I-mutant, DynaMut2, CUPSAT, mCSM-PPI2, and mCSM were used. By utilizing these different methods, a consensus can be reached on the calculation of this change in binding affinity. All methods predicted that the NSP14/NSP10 complex would be destabilized because of this mutation, with  $\Delta\Delta G$  (difference in Gibbs free energy) ranging between  $-1.12$  and  $-3.32$  kcal/mol ([supplementary table S5, Supplementary Material online](#)). This change in affinity may disrupt communication networks across the complex, resulting in a



**FIG. 4.** The interaction of the PHE60 of NSP14 in the wild-type (A) structure and SER60 in the mutant structure (B) and their respective interaction networks. Types of interactions are labeled, with the color of the dotted lines depicting the frequency of those interactions averaged over the MD simulations (VDW, Van der Waals interactions; HBONDS, hydrogen bonds; PIPSTACK, Pi–Pi stacking; IONIC, ionic bond; PICATION, Pi–cation interaction).



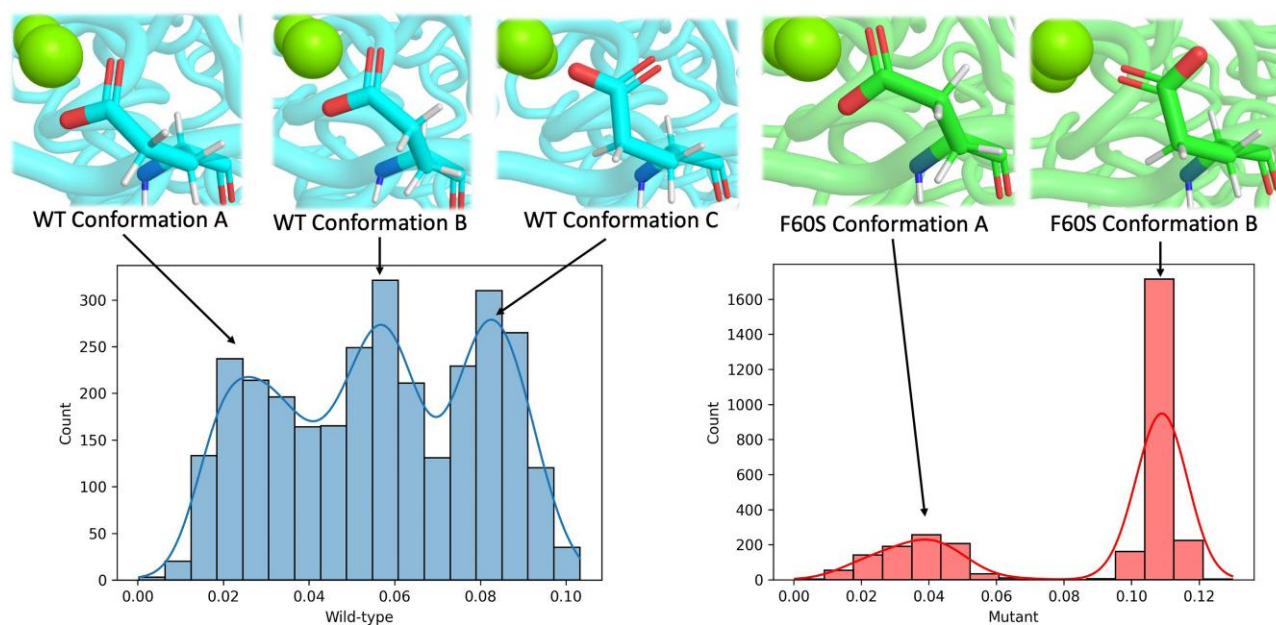
**FIG. 5.** (A) Eigenvector centralities projected onto wild-type (bottom) and mutant (top) NSP14/NSP10 complexes. Here, the centrality of each residue is depicted by color, with blue showing low centrality and red showing high centrality. CC depicts the catalytic core, whereas AS identifies a possible allosteric site. The silver sphere is the F60S mutation site. The green line is the between NSP14 and NSP10. (B) Change in contact frequencies between residues within 4.5 Å of catalytic residues.

disruption of allosteric signaling. The interface surface size was calculated using the Protein Interfaces, Surfaces and Assemblies (PISA) tool (Krissinel and Henrick 2007) which revealed a surface interface size of 2,256 Å<sup>2</sup> and energy of −24.3 kcal/mol.

### Mapping of Eigenvector Centrality and Changes in Catalytic Residue Interactions to Illuminate Allosteric Effects

Identifying the pathways responsible for allosteric communication in biomolecules is a difficult task, partly due to the





**Fig. 6.** Comparison of the rotamer conformation distributions in the wild-type (left) and F60S mutant (right) structures, with representative frames from MD trajectory (top).

intricate nature of these systems and the absence of effective characterization methods. However, this was previously achieved through exploiting eigenvector centrality, which in essence is a measure of the importance of, or contribution of, information from a node (in this instance a residue) to the remainder of the network (Negre et al. 2018). From MD simulations, nodes and edges can be derived from residues and interactions with well-connected/interacting nodes associated with high eigenvector centralities. Changes in these centralities can provide insight to system-wide changes which may have been brought about by mutations. Here, the *correlationplus* software (Tekpinar et al. 2021) was used to perform this analysis and identify eigenvector centralities within the wild-type and mutant NSP14 structures. Results for the wild-type show a focused eigenvector centrality within the NSP10 (fig. 5A), with weaker signals being detected at the catalytic core of the NSP14. However, within the mutant structure, the eigenvector centrality is now focused within the catalytic core of the NSP14, evidencing a reverse in eigenvector centrality. The shift in eigenvector centrality from the middle of the NSP10 protein to the catalytic core in the mutant structure suggests the interaction between NSP10 and NSP14 has altered in such way that affects the connectivity of distal regions of the complex. Differences were also observed between the wild-type and mutant structures in terms of residue contact frequencies at the catalytic core (fig. 5B). Here, we show that numerous residue–residue interactions throughout the catalytic core had changed in frequency, with increased interactions between E92, D273, and H268 and reduced interactions between E92, L149, and F146, further exemplifying changes at the network level within the system.

### Characterizing Structural Changes to Catalytic Residues

Considering the intensification of eigenvector centrality within the catalytic core of NSP14 and changes to the residue interaction network, we next sought to determine whether any structural changes that might impede function could be elucidated by further analyzing the MD data. It has previously been shown that the stability and flexibility of catalytic residues can be characterized through RMSD analysis, which can then show the population distribution of rotamer conformations (Riziotis et al. 2022). This investigation revealed that the distribution of catalytic residue conformations for D90, E191, D273, and H268 was similar between the wild-type and mutant structures across the  $3 \times 100$  ns MD simulations (supplementary fig. 5, Supplementary Material online). However, the distribution of rotamer conformations for E92 in the mutant structure showed a significant departure from the wild-type distribution (fig. 6). The wild-type showed three distinct populations of conformations (wild-type conformations A, B, and C), but analysis showed that E92 of the F60S mutant structure primarily resided in a single conformation (F60S conformation B) with very brief periods in conformation A. Conformation A of the F60S simulation is comparable with conformation B of the wild-type simulation; however, conformation B of the F60S simulation is unique and not observed in the wild-type simulation, which sees the carboxyl group of E92 rotate 180 degrees. This may indicate that the intrinsic flexible and dynamic quality of the wild-type E92 is no longer present within the mutant structure.

### Discussion

The exonuclease within the NSP14/NSP10 complex plays an important role in maintaining the accuracy of

replication by removing erroneous nucleotides introduced by RdRp. Examining the global SARS-CoV-2 genomic data set, we were able to identify clusters of cases which shared a mutation in NSP14 exonuclease that exhibited a detectable increase in evolutionary rate compared with other contemporaneous lineages.

The formation of the NSP10/NSP14 complex is known to promote exonuclease activity (Bouvet et al. 2014) and plays a key role in maintaining fidelity during viral RNA replication. The NSP10/NSP14 interface is relatively extended covering  $\sim 2,250 \text{ \AA}^2$  and a predicted interface energy of  $-24.3 \text{ kcal/mol}$ . Thus, a single mutation like F60S with a predicted interface energy loss of  $-1.1$  to  $-3.2 \text{ kcal/mol}$  is unlikely to lead to total complex disassembly.

The F60S mutation is anticipated to influence the exonuclease activity and evolutionary pace of the SARS-CoV-2 virus via an allosteric mechanism triggered by modifications in the interaction network within the NSP14/NSP10 interface. Although it is probable that this complex will stay structurally intact, the mutation is expected to modify the surface area and perturb the allosteric communication between the interface of the complex and the active site of NSP14. Figure 1 provides a visual representation of the active site's positioning in relation to the NSP14/NSP10 interface.

Our modeling suggests that the NSP14 F60S mutation results in a significant reduction in the number of residue-residue interactions within itself and NSP10. Although the latter is to be expected given its site at the interface and this is borne out in a slight reduction in predicted affinity, the reduction in interactions within NSP14 is likely to be more significant due to the extensive number of reduced contacts. Additionally, there was a notable transference in eigenvector centrality following the introduction of the F60S mutation. This became focused within the catalytic core of the mutant NSP14 as opposed to the wild-type NSP10, signifying a change in the epicenter of the communication network across the complex. Furthermore, an altered coordination of the catalytic core of the F60S mutant was characterized by the dramatic change in the distribution of rotamer conformations of the critical catalytic residue E92.

Simulations of the wild-type structure showed three distinct populations of conformations for E92, whereas the mutant structure primarily resided in a single conformation with only a brief period in one other conformation. This departure from the wild-type conformation distribution suggests that the intrinsic flexibility and dynamic quality of the wild-type E92 is no longer present in the mutant structure.

The change in the distribution of rotamer conformations for E92 of the catalytic core within the NSP14 would likely impact the function of the exonuclease, as one of the key roles of E92 is the accurate coordination of an  $\text{Mg}^{2+}$  ion, essential for the exonucleolytic reaction (Hwang et al. 2018). The catalytic core plays a critical role in the RNA hydrolysis process, and any changes to its stability

and flexibility could have consequences on the efficiency of its exonucleolytic reaction.

Mutations in the catalytic core of NSP14 have previously been linked to an increased accumulation of mutations in certain coronaviruses (Eckerle et al 2010; Graepel et al. 2017; Niu et al. 2021). Similarly, our study has demonstrated that the F60S mutant displays an increase in the rate with which it acquires mutations, a potential indication of a nonoptimal exonuclease.

The NSP14/NSP10 complex's function may be significantly impacted by the decrease in the number of residue interactions and binding strength at the interface, as well as the suggestive shift in the allosteric network and changes in rotamer conformation. If the binding affinity and/or the coupled residue interaction network between NSP14 and NSP10 decrease, the complex's allosteric communication networks could be disrupted, leading to a disruption of the exonuclease's proper functioning.

Moreover, it's worth noting that the observed evolutionary rates in this study are quite remarkable, surpassing those documented in other research efforts identifying lineages with heightened evolutionary rates (Hill et al. 2022; Tay et al. 2022). Although the mechanisms leading to an increased rate varies between these studies and the analysis of the F60S mutant in question, it is probable that selective host pressures exert an influence in all cases.

Here, we suggest that the F60S lineage experienced mutational meltdown, leading to its extinction within 22 days of detection. However, prior to its demise, the high mutation rate allowed these viruses to explore the sequence space and acquire mutations associated with improved fitness and increased virulence. Through an analysis using USHER (Turakhia et al. 2021) (supplementary table S7, Supplementary Material online). We identified several spike protein mutations, including 11 nonsynonymous mutations known to evade immune responses, improve receptor binding, or enhance transmissibility. Forty-four other nonsynonymous mutations were also detected within this lineage; though their significance remains uncertain, they could be of no effect or detrimental. We suggest that the F60S mutant eventually crossed a fitness threshold where the negative impact of detrimental mutations surpassed the benefits gained from advantageous mutations, resulting in cessation due to mutational meltdown.

In conclusion, this study has leveraged phylodynamic investigations to identify altered mutational rates within specific genotypes of SARS-CoV-2, which were then structurally assessed using MD approaches. Having identified F60S as a mutation of interest, our analysis then highlighted the impact of the NSP14 F60S mutation on the structural and functional integrity of the NSP14/NSP10 complex in SARS-CoV-2. The reduction in binding strength, shift in allosteric network, and changes in rotamer conformation could result in a reduction in the efficiency of the exonucleolytic reaction and negatively impact the replication fidelity of the virus, potentially providing an explanation for the increased evolutionary rate observed through the Bayesian analysis.



The mechanisms by which SARS-CoV-2 and other *Coronaviridae* generate diversity remain an important puzzle to decipher. The ability of SARS-CoV-2 to shuffle diversity via recombination has already been shown (Jackson et al. 2021), whereas the role of chronic carriers has also been identified as one mechanism by which SARS-CoV-2 may generate significant amounts of diversity, allowing “jumps” between variants (Hill et al. 2022). Furthermore, zoonotic (reverse zoonotic) SARS-CoV-2 infections in white-tailed deer (Pickering et al. 2022) and mink (Porter et al. 2023) have also been associated with increased evolutionary rates of  $3.7 \times 10^{-3}$  and  $6.59 \times 10^{-3}$  respectively, most likely as a result of significant selective pressures brought about by novel host environments.

This work potentially identifies another route by which mutations can be generated rapidly, which, when combined with the ability of SARS-CoV-2 to recombine, may be of significance in the evolution of SARS-CoV-2 going forward. It also demonstrates an interesting and potentially important observation, but further studies are now needed to fully understand the impact of exonuclease mutations and their potential impact on the generation of diversity in SARS-CoV-2 and other viruses.

## Methods

### Whole Genome Sequences Retrieval and Preparation

A search of the GISAID database was undertaken in November 2022 to identify clusters of SARS-CoV-2 viruses with mutations at the NSP14/NSP10 interface which were previously highlighted. Sequences with low coverage and missing collection date were excluded. Sequences were downloaded and allocated a data set only when >20 sequences could be attributed to a single PANGO lineage (O’Toole 2021). Each experimental data set was named in accordance with the mutation possessed at the NSP14/NSP10 interface. Two control lineages were also obtained from GISAID which did not contain mutations in either the catalytic core of the exonuclease or the NSP14/NSP10 interface. Each data set was checked for monophylogeny using USHER (Turakhia et al. 2021) to ensure close clustering with outliers and genetically distant sequences being removed. A list of Spike protein mutations was also obtained from USHER (supplementary table S7, Supplementary Material online). Each data set was aligned to the SARS-CoV-2 reference genome MN908947.3 using MAFFT (Kato and Standley 2013) with `--keeplength --6merpair --addfragments` options. Highly variable ends and troublesome sites were masked as described elsewhere (De Maio et al. 2020). The GISAID accession numbers for all sequence are available in supplementary materials.

### Temporal Signal and Bayesian Analysis

Maximum likelihood phylogenies were constructed with IQ-TREE (Trifinopoulos et al. 2016) using the GTR model (Tavaré 1986). A temporal signal was determined using

TempEst v.1.5.3 (Rambaut et al. 2016) with  $R^2$  and correlation coefficients being generated under the heuristic residual mean squared function with a best fitting root applied. Evolutionary analysis was undertaken to determine the number of *s/s/y* using BEAST v.1.10.4 (Suchard et al. 2018). Three different models were evaluated for each group. These models were composed of three different clock models: strict, lognormal uncorrelated relaxed, and exponential uncorrelated relaxed (Drummond et al. 2006). In each model, an exponential coalescent tree prior was implemented (Drummond et al. 2002) which is not expected to have an excessive impact on substitution rates while allowing the models to be more comparable (Ritchie et al. 2017). Each model also used a GTR+Γ4 substitution model. Population size was set to an exponential distribution with a mean of  $10^5$ , whereas for the growth rate, a Laplace distribution with a location of 0 and a scale of 100 was chosen. For the lognormal relaxed clock models, an exponential prior of 0.33 (mean) was set. A continuous-time Markov chain (CTMC) prior was set as the clock rate for each model (Ferreira and Suchard 2008). Each model was run in triplicate for chain lengths of  $10^8$ , sampling every  $10^3$  iterations. The best model was identified following the comparison of log-marginal likelihood estimates generated by SS analysis (Baele et al. 2012, 2013). The SS analysis was performed for 200 steps with chain lengths being set to  $10^6$ , and values were logged at every 1,000th iteration. 10% of results were discarded and considered burn in. Triplicate runs were combined using LogCombiner and Markov chain Monte Carlo (MCMC) traces were reviewed using Tracer v1.7.1 (Rambaut et al. 2018) to check for convergence and that effective sample sizes were >200. The best model was reran for each data set without tip-dates as a formal test of temporal signal (Duchene et al. 2020), and log-marginal likelihoods of data sets containing tip-dates and those without were then directly compared. Where a Bayes factor <1 in favor of the tip-date containing data set was reported, that data set was regarded as having a weak temporal signal and was removed from the study as inference would be unreliable. The data sets that were removed were H26M, M57I, M57V, and M62V (full BETS results are provided in supplementary table S6, Supplementary Material online). Finally, for those data sets with a strong temporal signal, an additional BEAST analysis was performed, sampling only from the prior. The results for this analysis are presented in supplementary figure 6, Supplementary Material online. Considering difference between the posterior densities of the prior-only and fully informed runs, it can be assumed that the results are not solely derived from the influence of the predetermined priors.

### Detection of Recombination

In May 2022, a globally representative data set of ~3,000 SARS-CoV-2 whole genome sequences was obtained from GISAID. This data set was downloaded from the Genomic Epidemiology Portal which houses the

preselected sequences which were considered being globally representative. Once downloaded, the ~3,000 sequence data set was separated to 10 tranches of ~300 sequences for computational efficiency. To each subset, the F60S mutant sequences were added, and as done previously, each tranche was aligned to the SARS-CoV-2 reference genome MN908947.3 using MAFFT (Katoch and Standley 2013) with `--keeplength --6merpair --addfragments` options. Again, highly variable ends and troublesome sites were masked as described elsewhere (De Maio et al. 2020). This produced 10 tranches of ~300 sequences of globally representative sequences combined with the F60S sequences. Each tranche was tested for recombination using RDP5 (Martin et al. 2021) with the following methods RDP, GENECOV, BootScan, MaxChi, Chimera, SiScan, PhyloPro, LARD, and 3Seq. All methods were used using default settings. The GISAID accession numbers for the sequences from the recombinant positive tranche are provided in [supplementary materials](#).

### NSP14/NSP10 Interface Analysis

The 7NOC structure of the SARS-CoV-2 NSP14 bound to NSP10, and RNA (Liu et al. 2021) was downloaded from the protein databank (<https://www.rcsb.org/>). The structure was uploaded to the RING3.0 server (Clementel et al. 2022), and using the default parameters for bond distances, an interaction map was generated. The resulting JSON file was visualized and analyzed in Cytoscape (Shannon et al. 2003).

### MD Analysis and RMSF Analysis

Again, using the 7NOC structure, two molecular systems were prepared for MD simulations with the CHARMM-GUI web interface (Jo et al. 2008). Except for introducing the F60S mutations to the mutant structure as mentioned previously, both systems were set up and ran in the same manner. A water box was made that was 10 Å from the protein edge, and 0.15 M of Na<sup>+</sup> and Cl<sup>-</sup> ions were added to neutralize the system charge. Energy terms were derived from CHARMM36 forcefields (Huang and Mackerell 2013), whereas water molecules were represented by TIP3 potentials (Jorgensen et al. 1983). A step size of 0.001 with a 12 Å cutoff was selected to consider nonbonded, distant interactions with a particle mesh Ewald summation (Darden et al. 1993). Hydrogen within covalent bonds were constrained by LINCS, and the system temperature was maintained at 310.15 K (37 °C). System pressure was maintained at 1 bar with a compressibility value of  $4.5 \times 10^{-5}$  bar<sup>-1</sup>. Equilibration was performed with a step size of 0.001 ns for 125,000 steps. Minimization was performed with a steep descent integrator for 5,000 steps. The production runs were performed at 0.002 ns step sizes for 50,000,000 steps (100 ns). All simulations were ran using GROMACS 2021.2 (Abraham et al. 2015). RMSD analysis was performed using the `gmx rms` command with the residue being selected as a reference from a custom index file.

RMSF analysis was performed with `gmx rmsf` using the carbon alpha of the protein system.

### Residue Interaction and Contact Frequency Analysis

The  $3 \times 100$  ns simulations for the wild-type were concatenated as were the simulations for the mutant. One thousand frames were then extracted from each of the concatenated simulations as PDB files. The files were then analyzed with RING3.0 (Clementel et al. 2022) locally. The resulting JSON file was visualized and analyzed in Cytoscape (Shannon et al. 2003). The concatenated PDB files were loaded to Cytoscape where SenseNet (Schneider and Antes 2022) was used to compare the change in contact frequency of residues within 4.5 Å of the catalytic residues.

### Binding Affinity Analysis

Five different servers were used to determine the change in binding affinity: I-mutant (<https://folding.biofold.org/i-mutant/i-mutant2.0.html>) (Capriotti et al. 2005), DynaMut2 (<https://biosig.lab.uq.edu.au/dynamut2/>) (Rodrigues et al. 2021), CUPSAT (<http://cupsat.tu-bs.de/>) (Parthiban et al. 2006), mCSM ([https://biosig.lab.uq.edu.au/mcsm/protein\\_protein](https://biosig.lab.uq.edu.au/mcsm/protein_protein)) (Pires et al. 2014), and mCSM-PPI2 ([https://biosig.lab.uq.edu.au/mcsm\\_ppi2/](https://biosig.lab.uq.edu.au/mcsm_ppi2/)) (Rodrigues et al. 2019)

### Calculating Eigenvector Centrality

Eigenvector centralities were calculated using the *correlationplus* software (Tekpinar et al. 2021). As described for the residue interaction analysis, the concatenated PDBs for both the wild-type and mutant structures were used as inputs for this analysis. Here, the dynamic cross correlation (DCC) approach was used with default parameters.

### Structure Illustrations

Structures were illustrated using ChimeraX (Goddard et al. 2018) and Pymol (Schrödinger and DeLano 2020).

### Supplementary material

Supplementary data are available at *Molecular Biology and Evolution* online.

### Acknowledgements

We thank the researchers who have shared genome data through the Global Initiative on Sharing All Influenza Data (GISAID). A.H.M. was supported by a studentship through the UKRI as part of the Artificial Intelligence, Machine Learning and Advanced Computing Centre for Doctoral Training (AIMLAC CDT) (grant code EP/S023992/1, project reference 2414958). Computational resources were provided by Advanced Research Computing at Cardiff (ARCCA) and Medical Research Council Cloud Infrastructure for Microbial Bioinformatics (MRC CLIMB) (grant code MR/T030062/1). We further acknowledge

resources provided through COVID-19 Genomics UK Consortium (COG-UK), which is supported by funding from the Medical Research Council (MRC) part of UKRI (grant code: MC\_PC\_19027). For the purpose of open access, the author has applied a CC BY public copyright licence (where permitted by UKRI, “Open Government Licence” or “CC BY-ND public copyright licence” may be stated instead) to any author accepted manuscript version arising. We extend our appreciation to Dr Zack Saud and Professor Richard Stanton for their initial guidance and resources for experimental work.

## Author Contributions

A.H.M. conceived the idea, sourced the data, managed the project, analyzed the data, and wrote the manuscript. G.M. contributed to the analysis and preparation of the MD simulations, MD analysis, and writing of the manuscript. A.S. contributed to the evolutionary analysis and writing of the manuscript. D.D.J. contributed to the analysis of the MD and structural data and writing of the manuscript. T.R.C. contributed to the conceptualization and analysis of the phylogenetic approach and preparation of the manuscript.

## Data Availability

All genomic data used in this analysis are available through GISAID. Accession identifiers are provided in [supplementary data](#), [Supplementary Material](#) online. All other data are available upon request.

## References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. 2015. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**:19–25.

Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Mol Biol Evol.* **29**(9):2157–2167.

Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol.* **30**(2):239–243.

Berger MB, Cisneros GA. 2023. Distal mutations in the  $\beta$ -clamp of DNA polymerase III\* disrupt DNA orientation and affect exonuclease activity. *J Am Chem Soc.* **145**(6):3478–3490.

Bouvet M, Imbert I, Subissi L, Gluais L, Canard B, Decroly E. 2012. RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc Natl Acad Sci USA.* **109**(24):9372–9377.

Bouvet M, Lugari A, Posthuma CC, Zevenhoven JC, Bernard S, Betzi S, Imbert I, Canard B, Guillemot JC, Lécine P, et al. 2014. Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem.* **289**(37):25783–25796.

Capirotti E, Fariselli P, Casadio R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**(suppl\_2):W306–W310.

Clementel D, Del Conte A, Monzon AM, Camagni GF, Minervini G, Piovesan D, Tosatto SC. 2022. RING 3.0: fast generation of

probabilistic residue interaction networks from structural ensembles. *Nucleic Acids Res.* **50**(W1):W651–W656.D.

Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. *J Chem Phys.* **98**(12):10089–10092.

De Maio, N., Walker, C., Borges, R., Weiglun, L., Slodkovicz, G. and Goldman, N., 2020. Issues with SARS-CoV-2 sequencing data. <https://virological.org/t/masking-strategies-for-sars-cov-2-alignments/480>

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**(5):e88.

Drummond AJ, Nicholls GK, Rodrigo AG, Solomon W. 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **161**(3):1307–1320.

Duchene S, Lemey P, Stadler T, Ho SY, Duchene DA, Dhanasekaran V, Baele G. 2020. Bayesian evaluation of temporal signal in measurably evolving populations. *Mol Biol Evol.* **37**(11):3363–3379.

Duffy S, Shackelton LA, Holmes EC. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet.* **9**(4):267–276.

Eckerle LD, Becker MM, Halpin RA, Li K, Venter E, Lu X, Scherbakova S, Graham RL, Baric RS, Stockwell TB, et al. 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.* **6**(5):e1000896.

Eckerle LD, Lu X, Sperry SM, Choi L, Denison MR. 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol.* **81**(22):12135–12144.

Eskier D, Suner A, Oktay Y, Karakulah G. 2020. Mutations of SARS-CoV-2 nsp14 exhibit strong association with increased genome-wide mutation load. *PeerJ.* **8**:e10181.

Ferreira MA, Suchard MA. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. *Can J Stat.* **36**(3):355–368.

Ghafari M, du Plessis L, Raghwanji J, Bhatt S, Xu B, Pybus OG, Katzourakis A. 2022. Purifying selection determines the short-term time dependency of evolutionary rates in SARS-CoV-2 and pH1N1 influenza. *Mol Biol Evol.* **39**(2):msac009.

Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, Morris JH, Ferrin TE. 2018. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**(1):14–25.

Gorbalenya AE, Enjuanes L, Ziebuhr J, Snijder EJ. 2006. Nidovirales: evolving the largest RNA virus genome. *Virus Res.* **117**(1):17–37.

Graepel KW, Lu X, Case JB, Sexton NR, Smith EC, Denison MR. 2017. Proofreading-deficient coronaviruses adapt for increased fitness over long-term passage without reversion of exoribonuclease-inactivating mutations. *MBio.* **8**(6):e01503–e01517.

Gribble J, Stevens LJ, Agostini ML, Anderson-Daniels J, Chappell JD, Lu X, Pruijssers AJ, Routh AL, Denison MR. 2021. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLoS Pathog.* **17**(1):e1009226.

Hill V, Du Plessis L, Peacock TP, Aggarwal D, Colquhoun R, Carabelli AM, Ellaby N, Gallagher E, Groves N, Jackson B, et al. 2022. The origins and molecular evolution of SARS-CoV-2 lineage B. 1.1.7 in the UK. *Virus Evol.* **8**(2):veac080.

Huang J, MacKerell Jr AD. 2013. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem.* **34**(25):2135–2145.

Hwang W, Yoo J, Lee Y, Park S, Hoang PL, Cho H, Yu J, Hoa Vo TM, Shin M, Jin MS, et al. 2018. Dynamic coordination of two-metal-ions orchestrates  $\lambda$ -exonuclease catalysis. *Nat Commun.* **9**(1):4404.

Jackson B, Boni MF, Bull MJ, Collier A, Colquhoun RM, Darby AC, Haldenby S, Hill V, Lucaci A, McCrone JT, et al. 2021. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **184**(20):5179–5188.

Jo S, Kim T, Iyer VG, Im W. 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem.* **29**(11):1859–1865.



- Jo WK, Drosten C, Drexler JF. 2021. The evolutionary dynamics of endemic human coronaviruses. *Virus Evol.* **7**(1):veab020.
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* **79**(2):926–935.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* **30**(4):772–780.
- Krissinel E, Henrick K. 2007. Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* **372**(3):774–797.
- Lin S, Chen H, Chen Z, Yang F, Ye F, Zheng Y, Yang J, Lin X, Sun H, Wang L, et al. 2021. Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Res.* **49**(9):5382–5392.
- Liu C, Shi W, Becker ST, Schatz DG, Liu B, Yang Y. 2021. Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science* **373**(6559):1142–1146.
- Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, et al. 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* **395**(10224):565–574.
- Ma Y, Wu L, Shaw N, Gao Y, Wang J, Sun Y, Lou Z, Yan L, Zhang R, Rao Z. 2015. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *PNAS.* **112**(30):9436–9441.
- Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, Kelz Z, Kumar V, Murrell B. 2021. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* **7**(1):veaa087.
- Minskaia E, Hertzog T, Gorbalenya AE, Campanacci V, Cambillau C, Canard B, Ziebuhr J. 2006. Discovery of an RNA virus 3' → 5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci USA.* **103**(13):5108–5113.
- Negre CF, Morzan UN, Hendrickson HP, Pal R, Lisi GP, Loria JP, Rivalta I, Ho J, Batista VS. 2018. Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci USA.* **115**(52):E12201–E12208.
- Niu X, Kong F, Hou YJ, Wang Q. 2021. Crucial mutation in the exoribonuclease domain of nsp14 of PEDV leads to high genetic instability during viral replication. *Cell Biosci.* **11**(1):106.
- O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, et al. 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evol.* **7**(2):veab064.
- Ogando NS, Zevenhoven-Dobbe JC, van der Meer Y, Bredenbeek PJ, Posthuma CC, Snijder EJ. 2020. The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol.* **94**(23):e01246-20.
- Parthiban V, Gromiha MM, Schomburg D. 2006. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Res.* **34**(suppl\_2):W239–W242.
- Peck KM, Lauring AS. 2018. Complexities of viral mutation rates. *J Virol.* **92**(14):e01031-17.
- Pickering B, Lung O, Maguire F, Kruczkiewicz P, Kotwa JD, Buchanan T, Gagnier M, Guthrie JL, Jardine CM, Marchand-Austin A, et al. 2022. Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nat Microbiol.* **7**(12):2011–2024.
- Pires DE, Ascher DB, Blundell TL. 2014. mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics.* **30**(3):335–342.
- Porter AF, Purcell DF, Howden BP, Duchene S. 2023. Evolutionary rate of SARS-CoV-2 increases during zoonotic infection of farmed mink. *Virus Evol.* **9**(1):vead002.
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* **67**(5):901–904.
- Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**(1):vew007.
- Ritchie AM, Lo N, Ho SY. 2017. The impact of the tree prior on molecular dating of data sets containing a mixture of inter- and intraspecies sampling. *Syst Biol.* **66**(3):413–425.
- Riziotis IG, Ribeiro AJ, Borkakoti N, Thornton JM. 2022. Conformational variation in enzyme catalysis: a structural study on catalytic residues. *J Mol Biol.* **434**(7):167517.
- Rodrigues CH, Myung Y, Pires DE, Ascher DB. 2019. mCSM-PPI2: predicting the effects of mutations on protein–protein interactions. *Nucleic Acids Res.* **47**(W1):W338–W344.
- Rodrigues CH, Pires DE, Ascher DB. 2021. DynaMut2: assessing changes in stability and flexibility upon single and multiple point missense mutations. *Protein Sci.* **30**(1):60–69.
- Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. 2010. Viral mutation rates. *J Virol.* **84**(19):9733–9748.
- Saramago M, Bárria C, Costa VG, Souza CS, Viegas SC, Domingues S, Lousa D, Soares CM, Arraiano CM, Matos RG. 2021. New targets for drug design: importance of nsp14/nsp10 complex formation for the 3'-5' exoribonucleolytic activity on SARS-CoV-2. *FEBS J.* **288**(17):5130–5147.
- Schneider M, Antes I. 2022. SenseNet, a tool for analysis of protein structure networks obtained from molecular dynamics simulations. *PLoS One* **17**(3):e0265194.
- Schrödinger L, DeLano W. 2020. PyMOL, Available at: <http://www.pymol.org/pymol>.
- Shannon A, Selisko B, Le NTT, Huchting J, Touret F, Piorkowski G, Fattorini V, Ferron F, Decroly E, Meier C, et al. 2020. Rapid incorporation of favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat Commun.* **11**(1):4682.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**(11):2498–2504.
- Shu Y, McCauley J. 2017. GISAID: global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* **22**(13):30494.
- Smith EC, Case JB, Blanc H, Isakov O, Shomron N, Vignuzzi M, Denison MR. 2015. Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity. *J Virol.* **89**(12):6418–6426.
- Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**(1):vey016.
- Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect Math Life Sci (Am Math Soc).* **17**:57–86.
- Tay JH, Porter AF, Wirth W, Duchene S. 2022. The emergence of SARS-CoV-2 variants of concern is driven by acceleration of the substitution rate. *Mol Biol Evol.* **39**(2):msac013.
- Tekpinar M, Neron B, Delarue M. 2021. Extracting dynamical correlations and identifying key residues for allosteric communication in proteins by *correlationplus*. *J Chem Inf Model.* **61**(10):4832–4838.
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**(W1):W232–W235.
- Turakhia Y, Thornlow B, Hinrichs AS, De Maio N, Gozashti L, Lanfear R, Haussler D, Corbett-Detig R. 2021. Ultrafast Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat Genet.* **53**(6):809–816.
- Zuo Y, Deutscher MP. 2001. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* **29**(5):1017–1026.