Mutational hotspot in the SARS-CoV-2 Spike protein N-terminal domain conferring immune escape potential

Slawomir Kubik[1]*, Nils Arrigo[2]*, Jaume Bonet[2], Zhenyu Xu[2†]

[1] Data Science Department, SOPHiA GENETICS, Chemin des Mines 9, 1202 Geneva, Switzerland

[2] Data Science Department, SOPHiA GENETICS, Rue du Centre 172, 1025 Saint-Sulpice, Switzerland

* these authors contributed equally to the work

[†] corresponding author

Keywords: SARS-CoV-2 genome, coronavirus, Spike NTD, W152, viral evolution, neutralizing antibody, immune escape

ABSTRACT

Global efforts are being taken to monitor the evolution of SARS-CoV-2, aiming at early identification of mutations with the potential of increasing viral infectivity or virulence. We report a striking increase in the frequency of recruitment of diverse substitutions at a critical residue (W152), positioned in the N-terminal domain (NTD) of the Spike protein, observed repeatedly across independent phylogenetic and geographical contexts. We investigate the impact these mutations might have on the evasion of neutralizing antibodies. Finally, we uncover that NTD is a region exhibiting particularly high frequency of mutation recruitments, suggesting an evolutionary path on which the virus maintains optimal efficiency of ACE2 binding combined with the flexibility facilitating the immune escape.

INTRODUCTION

RNA viruses display particularly high mutation rates (1), with SARS-CoV-2 undergoing approximately 10^{-3} substitutions/site/year (2). Globally, the selective pressure imposes conservation of adaptive mutations facilitating the viral spread. The overall success of viral transmission depends on the mutation rate, the extent of immune response, and the population size (3). During the pandemic, where population size is large, rapid increase in the frequency of alterations is observed at critical positions of the viral genome. Two commonly reported forces shaping the natural selection for SARS-CoV-2 are the adaptation to host (4) and the evasion of the immune response (5), including immunity triggered by the vaccines (6). Consequently, the evolutionary rate is particularly high for the S gene encoding the Spike protein (7), the main contact point with the ACE2 receptor of the host cell (8). Importantly, Spike serves also as the immunizing agent in the majority of COVID-19 vaccines (9).

It is expected that mutations improving viral fitness emerge independently across unrelated viral clades. An example of an adaptive mutation that emerged relatively early during the pandemic is D614G substitution in Spike, by the end of 2020 present in almost every SARS-CoV-2 genome in the world (10) and believed to improve the Spike trimer interaction with ACE2 (4,11). Since the last months of 2020, increase in frequency of other mutations was observed, with the N501Y and E484K being two prominent examples. The mechanisms by which they confer evolutionary advantage to SARS-CoV-2 vary. Particularly, N501Y increases the adaptation to host by enhancing interaction with the ACE2 receptor (12–14) resulting in more efficient transmission (15). In contrast, E484K appears as selectively advantageous by decreasing the strength of interaction with neutralizing antibodies (5,16,17), which facilitates evasion of the immune response. More recently, L452R substitution was reported to have similar properties to E484K (5,16,18,19). Importantly, these mutations have arisen independently within diverse, unrelated genomic contexts, and at distant geographical locations, being examples of convergent evolution. Moreover, it may be expected that certain genomic positions under strong negative frequency-

dependent selection – as expected in the context of immunity-escaping processes (20) – will display a diverse spectrum of mutations.

Adaptive traits require close monitoring, particularly because they are likely to appear as increasingly prominent within SARS-CoV-2 strains under the current global vaccination efforts aiming at establishing herd immunity. Several studies focused on evaluating potential impact of mutations on the viral spread and antibody evasion (16,21–27). Most investigations focused on the receptor binding domain (RBD) of the Spike, the immunodominant part of the protein (28) containing the ACE2-interacting interface. However, mutations at sites outside of the RBD, such as D614, might also have strong impact on both, the infectivity and immune escape. For example, the N-terminal domain (NTD) of the Spike was shown to be a potent target for neutralizing antibodies (6,29,30).

By screening SARS-CoV-2 genome sequences for residues undergoing frequent and diverse mutations we pinpointed W152, a residue present in NTD, whose alterations have the potential of being advantageous for viral transmission. We identified that several substitutions, leading to a limited set of amino-acid changes at position W152, were independently recruited numerous times across many distantly related phylogenetic contexts and diverse geographical locations, suggesting their adaptive character. Insights from structural studies confirm that the identified W152 substitutions remove an important interaction point for multiple potent neutralizing antibodies. Furthermore, we demonstrate that mutations in NTD were recruited more frequently than in other regions of Spike during the second wave of the pandemic, likely due to improving viral fitness through the immune escape. Our work highlights the importance of monitoring individual mutations occurring outside of the Spike RBD.

METHODS

Identification of emerging mutations in DDM database

Aggregated statistics were collected from SOPHiA DDM database (https://www.sophiagenetics.com/technology/). These included list of mutations detected in SARS-CoV-2 genotypes at a fraction of at least 70% (31) and the country of origin. Clades were assigned to the genotypes using Nextclade (version 0.13.0) (32) and lineages using Pangolin (version 2.3.2) (33). We then filtered the dataset to keep distinct, recurring mutations at a single genomic position, present at least 5x more frequently in DDM than globally (based on data deposited in GISAID, https://www.gisaid.org), having emerged across multiple clades and lineages.

Analysis of publicly available SARS-CoV-2 sequences

Phylogenetic and protein data

We extended our analysis to publicly available data included in the Audacity global COVID phylogeny along with all Spike protein sequences (1'028'876 entries - spikeprot0406.fasta) deposited in GISAID as of 2021/04/11. Protein sequences were aligned against the Spike reference (YP_009724390.1 as obtained from NCBI, as of 2021/04/11), using muscle v3.8.31 (34) with default parameters for protein analysis and converted into VCF files using custom R scripts (35). The Audacity phylogeny and VCF files were then merged to obtain a phylogeny of 566,422 tips (391,504 internal nodes) with Spike protein information available for all tips.

Identification of independent W152 recruitments

Our analysis aimed at inventorying independent recruitments of Spike mutations. From a phylogenetic standpoint, the task required to regroup SARS-CoV-2 genomes holding a Spike mutation of interest into sets of genomes that shared a common ancestor (i.e. "clades"). Assuming rare recombination, the most recent common ancestor (i.e. "mrca") of a clade marked the "recruitment event" at which the mutation of interest arose in the tree. The remainder of the clade then replayed transmission of the mutant to new hosts and the creation of a contagion cluster.

Because the size of the Audacity tree rendered most ancestral character reconstructions intractable, we opted for an ad-hoc heuristic that iteratively delineated clades and identified the respective mrca of mutation carrying sequences. To this end, we applied a tree walk algorithm that identified clades given a tree topology and a set of tips states. Our heuristic proceeded as follows (see **Fig. S1A**): (i) identify all tree tips containing a mutation of interest (i.e. "mutant tips"), (ii) for each mutant tip; move up to the parent node and collect all of its descending tips, (iii) calculate the proportion of retrieved tips that bear the mutation of interest (tips with missing data and tips already visited by the heuristic were excluded from the computation), (iv) if the proportion exceeded 0.95: accept the parent node so as to increment the focal clade with a set of genomes that consistently contain the mutation of interest – while still allowing for punctual reversions to WT genotypes- and repeat steps (ii) to (iv). Keep ascending the tree until reaching a parent node where the threshold is not met. Once stopped, define the last visited node as the mrca and gather all descendant tips. The tree walk was initiated from every mutant tip and returned a collection of clades that was finally refined by merging clades that shared at least one genome.

The independent clades of W152 mutations were then characterized according to: amino acid change, clade size (number of genome occurrences belonging to the clade), earliest observation (deposition week, taken as a proxy of recruitment time), geographical spread (number of countries in which a clade was found) and status of concerning RBD mutations (i.e. L452R, E484K, N501Y) at the recruitment event (inferred by checking for the presence of RBD mutations shared within genomes closely related to the clade of interest). Finally, the analysis was extended to all Spike positions to establish the overall distribution of recruitment events of Spike mutations, as a function of time. The R packages ape v5.0 and fishplot v.0.5.1 were used to perform all phylogenetic analyses. All scripts were deposited on github.

Structure-based prediction of the impact of the mutations on Spike interactions

We assessed the effect of W152 mutants on neutralizing antibody (nAb) recognition by generating single point mutants (W152C, W152L and W152R) and evaluating their changes in binding free energy (ddG) against 5 different NTD-target antibodies (1-87 and 5-24 (36), 4A8 (29), FC05 (37) and S2X333 (38),

structures obtained from the Protein Data Bank (39)) using Rosetta (40). nAbs were selected done based on the availability and interaction angle (36) to provide a broad view of the possible scenarios. For each experiment (mutant-antibody pair), side chain minimization was performed after the mutation and before the ddG analysis. As minimization in Rosetta is a stochastic-based process, a total of 100 decoys were generated for each experiment to define a distribution of ddG values. Finally, all decoys of ddG (regardless of the mutation) for a given antibody were normalized to the distribution obtained with wild-type (WT) Spike for that antibody.

RESULTS

Tryptophane at position 152 is a mutational hotspot

We screened SARS-CoV-2 genomes present in the DDM database (see Methods section) in order to identify novel, potentially concerning mutations within the S gene, defined as (i) multiple non-synonymous substitutions present at a single position (ii) displaying increased frequency in comparison with global frequency, (iii) independent recruitments across multiple lineages and (iv) across multiple geographical locations. This approach identified distinct mutations at position W152 of the Spike NTD resulting in substitution of tryptophane to leucine (W152L) or arginine (W152R) (**Table 1**). Moreover, two other, less frequent substitutions at the same position were identified in DDM: cysteine (W152C) or glycine (W152G). Importantly, both W152L and W152R were reported across unrelated phylogenetic contexts (i.e. clades and lineages) and at distant geographical locations. Of note, W152C is a clade-defining mutation for the CAL.20C clade (B.1.429 lineage) responsible for an outbreak in Southern California (41). Frequent and diverse mutations at a single residue, emerging repeatedly and independently, suggested their putative adaptive role.

Diverse W152 mutations were recruited independently in multiple phylogenetic contexts

We investigated the recruitment dynamics of W152 mutations in global datasets deposited in GISAID database. Due to a relatively low number of depositions in weeks 76-80 of the pandemic comparing to the preceding weeks (**Figure S1B**) we only took into account depositions with collection date up to week 75. Rapid and steady growth was observed in the number of submitted sequences bearing three W152 substitutions (W152C, W152L and W152R) during the second wave of the pandemic, in the period between December 2020 and April 2021 (weeks 56-75 of the pandemic) (**Figure 1A** and **1B**).

These substitutions were associated with 171 independent recruitments since the beginning of the pandemic (**Figure 1C**), with only sporadic cases (14.6%, 25/171) reported during the first wave (until week 55 - **Figure 1D**). The largest cluster was reported for W152C with over 13'000 occurrences in 30 countries (including the CAL.20C lineage, referred to as the "California variant"), with the second-largest containing almost 1'500 sequences bearing W152L and present in 20 countries. However, most clusters were relatively small in size (≤5 sequences reported for 86% clades [147/171]) and present in only 1 country (**Figure S1C**). This observation pointed to frequent and independent recruitments rather than spreading of viruses bearing W152 mutations due to cross-border transmission. During weeks 65-75, the number of independent W152 mutation recruitments was in the range between 90th and 99th percentile among independent mutation recruitments reported for all Spike positions (**Figure 1E**). These observations placed W152 as one of the most dynamic Spike positions in terms of mutation recruitments, just behind N501, E484, and ahead of L452. W152 was also one of only two Spike residues with 3 independent substitutions having at least 500 occurrences in GISAID being reported for each, with another NTD residue (D80) being the other one (**Figure S1D**).

In order to investigate whether W152 substitutions might confer direct evolutionary advantage to SARS-CoV-2, we investigated Spike mutations most frequently co-occurring with each W152 substitution (**Figure 2A**). The only Spike mutation co-occurring with all three tested W152 substitutions was the ubiquitous D614G. Each substitution displayed relatively high frequency of different co-occurring Spike mutations. For W152C, the most frequently co-occurring Spike mutations were S13I and L452R, both

clade-defining for the 20C/S:452R clade (lineage B.1.429). In case of W152L these were E484K and G769V (found in R.1 lineage) and for W152R – HV69del, Y144del, N501Y, A570D, P681H, T716I, S982A and D1118H, defining for clade 20I/501Y.V1 (lineage B.1.1.7). In most clades (67.8%, 116/171) W152 substitution did not co-occur with any of the prominent RBD mutations suspected of being advantageous (**Figures 2B** and **2C**). However, the majority of sequences in each of the largest clusters for individual substitutions contained at least one of these: L452R (for W152C), E484K (W152L) or N501Y (W152R) (**Figure 2C**). After excluding the single largest cluster for each substitution, the fraction of sequences without adaptive RBD mutation was 64%. Importantly, our phylogenetic analysis indicated that the ancestor clades of each of the largest clusters were already bearing RBD mutations when acquiring W152 substitutions (see Methods section). The above observations point to a potential key role of the Tryptophane at position 152 for SARS-CoV-2 fitness, regardless of the co-existing adaptive RBD alterations.

W152 is an important interaction point for neutralizing antibodies

Distinct regions of the Spike are subject to different evolutionary constraints. RBD domain mutations can be steered by both: increased ACE2 binding and neutralizing antibody (nAb) evasion while the NTD alterations are mostly determined by the nAb escape potential. We sought to investigate how W152 mutations affect the recognition of the NTD domain by selected NTD-targeting nAbs. To this end, we introduced each of the mutations of interest (W152L, W152R and W152C) into 5 Spike NTD structures, each bound to a different NTD-nAb (see Methods section), performed side-chain optimization and ddG analysis (**Figure 3A**). In all cases we observed a residue in the Ab chain engaged in a pi stacking interaction with W152 in the wild-type Spike protein (**Figure 3B**). 1-87 and 4A8 nAbs wrap W152 inside one of the CDR loops, making it a key position contributing to the interaction. In 5-24 the position is located just outside of the interface, but close enough for the minimal conformational changes to allow it to participate in the binding. FC05 and S2X333 use W152 for secondary interactions, thus the

contribution of the residue to the interface is negligible and its mutation can be easily compensated by a side chain movement.

The stacking interactions were lost with any of the considered mutations, thus effectively decreasing the affinity of the nAb to bind to NTD (**Figure 3A**). As expected, the effect varied depending on the extent of the W152 participation in the main interacting surface. Thus, in the case of the nAbs 1-87 (**Figure 3C**) and 4A8 (**Figure 3D** a significant drop in affinity was observed, while in peripherically-interacting nAb such as 5-24, FC05 and S2X333 the effect was little or negligible (**Figure 3A**). In case of 1-87 and 4A8 the amino acid substitution not only results in the loss of the pi stack interaction but also affects the pocket generated by the antibody's CDR loop, which leads to a substantial loss of the binding affinity. The effects are especially drastic for W152L due to the exposure of the hydrophobic side chain.

NTD mutations are enriched in the second wave of the pandemic and point to immune escape potential

The W152 residue is placed in the vicinity of mutations of the NTD domain that received increasing attention owing to recent emergences in the variants of concern (**Figure 4A**, L18, H69 or Y144). NTD constitutes an exposed part of the Spike protomer, making it a prominent target for antibodies, yet, contrary to RBD, mutations within NTD have potentially little impact on receptor binding. We investigated the possibility that residues present in NTD undergo extensive mutagenesis facilitating immune escape without hampering the interaction with ACE2. During the first wave of the pandemic (until week 55) mutation recruitments were distributed relatively uniformly across Spike domains (**Figure 4B**). On the contrary, during the second wave (i.e. week 56 and onwards) NTD displayed an elevated number of mutation recruitments comparing to other parts of the protein (**Figure 4C**). This localized bias in diversity was statistically significant (**Figure 4D**) and could result from adaptive changes in response to global immunity. The number of mutation recruitments per position correlated significantly with the evolutionary lability of the Spike protein observed across *Coronaviridae* (**Figure S2**; $p<10^{-5}$ for each

tested region). Nevertheless, recruitment events were significantly more common for the NTD in comparison to RBD and the remainder of the Spike protein residues, confirming similar evolutionary pattern for the related viruses. These observations suggest that globally, NTD mutations have higher propensity of being advantageous than alterations in other regions of the protein.

DISCUSSION

Accumulating body of evidence suggests a key role of Spike NTD mutations in viral evolution. Based on conservation estimates and the number of independent mutation emergences we demonstrate that this domain undergoes more rapid evolutionary changes in comparison with other parts of the protein. This localized evolution gained momentum during the second wave of the pandemic, likely in response to global increase in immunity. The most prominent forces driving selection of Spike mutations are the increased interaction with ACE2 and the evasion of neutralizing antibodies (nAb). NTD and RBD constitute the most exposed parts of the Spike making them the most likely targets of the immune response. As demonstrated by us and others, mutations in these domains often facilitate immune evasion (5,6,16,17,21,24,25,36). However, RBD mutations are more evolutionarily constrained due to their role in interaction with ACE2. Given the relatively small contribution of NTD to ACE2 binding, alterations in this domain might constitute the evolutionary 'disguise' the virus uses to avoid antibody neutralization. The variability is not restricted to amino acid substitutions as a significant number of deletions was also reported in NTD and linked to the immune escape (42). Identification of potential nAb that can be used against different variants of RBD and NTD is of great importance, considering that antibody cocktails might act collaboratively to impede the progression of viral infection (37).

We identified W152 as a NTD residue undergoing particularly extensive evolutionary dynamics, highlighted by multiple individual substitutions emerging across many phylogenetic and geographical contexts. Remarkably frequent mutation recruitment events were reported at this position globally, with a clear increase in intensity since the end of 2020 (week 55 of the pandemic). The largest clusters reported

for each of the three frequent substitutions – W152L, W152R or W152C – were characterized by the co-recruitment of one of the prominent, adaptive RBD mutations (E484K, N501Y or L452R, respectively). Although the contribution of W152 mutations to those three particular events could not be decoupled from that of their co-occurring alterations in RBD, our results suggest an adaptive role of the W152 substitutions as most of their recruitment events did not occur in parallel with RBD mutations. In line with this finding, recent study demonstrated that W152C allows to further increase B.1.429 infectivity in comparison to L452R alone (18).

It is generally appreciated that advantageous mutations initially arise as the quasi-species, present only in a fraction of viral genomes within a given host. Providing a competitive edge, they are progressively increasing in prevalence and are eventually transmitted to new hosts giving rise to new clades responsible for infection clusters. In this regard, the advantage conferred by W152 mutations might be exemplified by a reported increase in the intra-host fraction of genomes bearing W152L substitution during the infection (43).

Tryptophane at position 152 was shown to have a role in stabilizing intermolecular interactions of the Spike (44). Our analysis suggests that mutations of W152 diminish the strength of interactions with several neutralizing antibodies, consistent with other reports (6,29,36). As the chemical properties of each side chain of the amino acids present in the mutants differ (leucine – hydrophobic, arginine – basic, cysteine – sulfide) the advantage likely stems from the fact of removing tryptophane (aromatic) from the relevant position rather than establishing specific novel interactions. Such alteration decreases the propensity of forming stable stacking interactions similar to those observed in the 1-87 or 4A8 antibody complexes with Spike (29). Alarming evidence supporting the role in W152 in evading immune response comes from a report of the R.1 lineage bearing W152L (45) being responsible for a local outbreak in a population having underwent a vaccination program (46).

Most SARS-CoV-2 evolution studies concentrate on the mutations present in RBD. Our work clearly suggests that large emphasis should be also put on monitoring alterations occurring in NTD. Dedicated

biochemical, immunological and cell biology investigations are necessary to understand the exact effects of the W152 substitutions on Spike properties as well as on SARS-CoV-2 infectivity or virulence. Significant efforts are spent on tracking the spread of specific variants of concern, such as the B.1.1.7 ("UK variant") or the B.1.167.2 ("Indian variant"). However, our study outlines the importance of monitoring the emergence, recruitment and spread of individual mutations as well. Mutations of residues such as W152, L18, H69, Y144, L452, E484 or N501 occurred frequently and have been recruited across many independent SARS-CoV-2 lineages. As of writing, the W152 mutations spurred 171 independent clusters that directly or indirectly contributed to contaminate over 15,000 patients. Hence, monitoring efforts should not overlook the fact that evolutionary advantageous mutations can emerge in virtually any SARS-CoV-2 lineage and at any geographical location.

TABLES

Table 1

mutation	DDM frequency [%]	Global frequency [%]	Country count	Continent count	Clade count	Clades	Lineage count	Lineages
Trp152Leu	0.82	0.06	3	2	5	19A, 20A, 20B, 20I/501Y.V1, 20H/501Y.V2	9	B, B.1, B.1.1, B.1.1.7, B.1.29, B.1.351, B.1.416, B.1.420, R.1
Trp152Arg	0.73	0.04	3	2	4	19A, 20A, 20D, 20I/501Y.V1	7	B, B.1, B.1.1.1, B.1.111, B.1.1.7, B.1.166, B.1.402
Trp152Cys	0.02	1.56	1	1	1	20C	1	B.1.429
Trp152Gly	0.02	2x10-4	1	1	1	20I/501Y.v1	1	B.1.1.7

Aggregated data on W152 mutations identified in DDM database

FIGURE LEGENDS

Figure 1

Diverse W152 mutations emerged independently across numerous genomic contexts. **(A)** Cumulative number of sequences bearing three W152 mutations uploaded to GISAID (y axis) depending on the upload date (x axis). **(B)** Number of sequences bearing three W152 mutations uploaded weekly to GISAID (y axis) depending on the upload date (x axis). **(C)** Representative phylogenetic tree of analyzed sequences (displaying 1% of the global Audacity tree); 171 independent W152C/L/R clades are identified in the present study and displayed to outline mutation recruitment events (dots) and ensuing contagion clusters (edges); largest cluster for W152C (CAL.20C) is indicated. **(D)** Fish plot of all SARS-CoV-2 genomes deposited in GISAID, with color-coded areas corresponding to the numeric abundance of clades that recruited W152C (purple – 20 events), W152L (turquoise – 75 events) or W152R (yellow – 76 events) mutations, as a function of the upload date. **(E)** Plot showing the weekly count of recruitments of clades bearing mutations at positions W152 (black), L452 (orange), E484 (red) and N501 (green); grey areas indicate values for percentiles 75th, 90th, 95th and 99th of mutation recruitments across all Spike positions.

Figure 2

Most W152 mutation recruitments are not associated with a known adaptive RBD mutation. (**A**) Plots showing, for the three W152 substitutions, the frequency of co-occurrence of non-synonymous Spike mutations per sequence, present at a frequency of at least 0.01; known, adaptive RBD mutations are indicated by orange, red and green labels. (**B**) As (A) but calculated per clade, shown for frequencies of at least 0.05. (**C**) Plot showing, for three W152 substitutions, independently recruited clades (circles)

colored depending on the co-existing adaptive RBD mutation and positioned depending on the clade size (x axis).

Figure 3

Interactions with neutralizing antibodies are weakened by W152 mutations. (**A**) Normalized ddG calculation with the wild-type Spike and the three W152 mutants (W152C, W152L, W152R) for binding to different anti-NTD nAb, based on available structures; for complexes with 1-87 and 4A8 W152 is found within the binding interface, for 5-24, FC05 and S2X333 complexes W152 is a secondary interaction point at the edge of the binding interface. (**B**) Zoomed-in view of W152-containing region for available structures of Spike in complex with anti-NTD nAbs; in all cases, the nAb presents a residue engaged in pi stacking with W152 (red box); aromatic amino acids are the binding partners for W152 in all nAbs except S2X333, where arginine is involved in the interaction. (**C, D**) Interaction interfaces of Spike W152 mutants with 1-87 and 4A8, respectively.

Figure 4

Extensive NTD evolution during the second wave of the pandemic. (A) Structural model of Spike trimer (PDB Id: 7DDD); RBD marked in green, NTD marked in red; chosen mutation positions are indicated. (**B**) Spike structure model colored according to the number of mutation recruitments for each amino acid position for weeks 1-55 of the pandemic; the color ramp partitions number of recruitments into eight equally spaced bins ranging between the minimum value (blue) and two times the average (light blue); all remaining observations that include more variable amino acid positions are color-coded in yellow. (**C**) As B but for weeks 56-75. (**D**) Boxplot showing mutation recruitment events for all Spike positions present in NTD (red), RBD (green) or the remainder of the protein (grey) for weeks 1-55 (left) or 56-75 (right); the statistical significance is assessed using permutation testing (10'000 permutations)

REFERENCES

1. Duffy S. Why are RNA virus mutation rates so damn high? PLOS Biol. 2018 Aug 13;16(8):e3000003.

2. van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Balloux F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat Commun. 2020 Dec;11(1):5986.

3. Seeholzer A, Frey E, Obermayer B. Periodic versus Intermittent Adaptive Cycles in Quasispecies Coevolution. Phys Rev Lett. 2014 Sep 15;113(12):128101.

4. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020 Aug;182(4):812-827.e19.

5. Greaney AJ, Starr TN, Gilchuk P, Zost SJ, Binshtein E, Loes AN, et al. Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition. Cell Host Microbe. 2021 Jan;29(1):44-57.e9.

6. McCallum M, Bassi J, Marco AD, Chen A, Walls AC, Iulio JD, et al. SARS-CoV-2 immune evasion by variant B.1.427/B.1.429 [Internet]. Immunology; 2021 Apr [cited 2021 Apr 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.03.31.437925

7. Vilar S, Isom DG. One Year of SARS-CoV-2: How Much Has the Virus Changed? Biology. 2021 Jan 26;10(2):91.

8. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020 Feb;395(10224):565–74.

9. Krammer F. SARS-CoV-2 vaccines in development. Nature. 2020 Oct 22;586(7830):516–27.

10. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science. 2020 Nov 12;eabe8499.

11. Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. Cell. 2020 Oct;183(3):739-751.e8.

12. Laffeber C, de Koning K, Kanaar R, Lebbink JH. Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants [Internet]. Biochemistry; 2021 Feb [cited 2021 Apr 13]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.02.22.432357

13. Luan B, Wang H, Huynh T. Enhanced binding of the N501Y□mutated SARS□CoV□2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. Bubb M, editor. FEBS Lett. 2021 Apr 3;1873-3468.14076.

14. Tian F, Tong B, Sun L, Shi S, Zheng B, Wang Z, et al. Mutation N501Y in RBD of Spike Protein Strengthens the Interaction between COVID-19 and its Receptor ACE2 [Internet]. Biophysics; 2021 Feb [cited 2021 Apr 13]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.02.14.431117

15. Liu Y, Liu J, Plante KS, Plante JA, Xie X, Zhang X, et al. The N501Y spike substitution enhances SARS-CoV-2 transmission [Internet]. Microbiology; 2021 Mar [cited 2021 Apr 13]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.03.08.434499

16. Starr TN, Greaney AJ, Dingens AS, Bloom JD. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Rep Med. 2021 Apr;100255.

17. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell. 2021 Feb;S0092867421002269.

18. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. Cell. 2021 Apr;S0092867421005055.

19. Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe. 2021 Mar;29(3):477-488.e4.

20. Brisson D. Negative Frequency-Dependent Selection Is Frequently Confounding. Front Ecol Evol. 2018 Feb 21;6:10.

21. Chen J, Gao K, Wang R, Wei G. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. ArXiv. 2020 Oct 13;

22. Garcia-Beltran WF, Lam EC, Denis KSt, Nitido AD, Garcia ZH, Hauser BM, et al. Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity [Internet]. Infectious Diseases (except HIV/AIDS); 2021 Feb [cited 2021 Apr 4]. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.02.14.21251704

23. Gobeil SM-C, Janowska K, McDowell S, Mansouri K, Parks R, Stalls V, et al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation and antigenicity [Internet]. Microbiology; 2021 Mar [cited 2021 Apr 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.03.11.435037

24. Liu Z, VanBlargan LA, Bloyet L-M, Rothlauf PW, Chen RE, Stumpf S, et al. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization [Internet]. Microbiology; 2020 Nov [cited 2021 Apr 4]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.11.06.372037

25. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. Science. 2021 Feb 19;371(6531):850–4.

26. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature [Internet]. 2021 Mar 8 [cited 2021 Apr 4]; Available from: http://www.nature.com/articles/s41586-021-03398-2

27. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. eLife. 2020 Oct 28;9:e61312.

28. Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, et al. Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. Cell. 2020 Nov;183(4):1024-1042.e21.

29. Chi X, Yan R, Zhang J, Zhang G, Zhang Y, Hao M, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science. 2020 Aug 7;369(6504):650–5.

30. Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature. 2020 Aug 20;584(7821):450–6.

31. Kubik S, Marques AC, Xing X, Silvery J, Bertelli C, De Maio F, et al. Recommendations for accurate genotyping of SARS-CoV-2 using amplicon-based sequencing of clinical samples. Clin Microbiol Infect. 2021 Apr;S1198743X21001646.

32. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Kelso J, editor. Bioinformatics. 2018 Dec 1;34(23):4121–3.

33. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Nov;5(11):1403–7.

34. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004 Mar 8;32(5):1792–7.

35. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/

36. Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, et al. Potent SARS-CoV-2 Neutralizing Antibodies Directed Against Spike N-Terminal Domain Target a Single Supersite [Internet]. Biochemistry; 2021 Jan [cited 2021 Apr 5]. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.01.10.426120

37. Wang N, Sun Y, Feng R, Wang Y, Guo Y, Zhang L, et al. Structure-based development of human antibody cocktails against SARS-CoV-2. Cell Res. 2021 Jan;31(1):101–3.

38. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 2021 Apr;184(9):2332-2347.e16.

39. Berman HM. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235–42.

40. Fleishman SJ, Leaver-Fay A, Corn JE, Strauch E-M, Khare SD, Koga N, et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. Uversky VN, editor. PLoS ONE. 2011 Jun 24;6(6):e20161.

41. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E. Emergence of a Novel SARS-CoV-2 Variant in Southern California. JAMA [Internet]. 2021 Feb 11 [cited 2021 Apr 4]; Available from: https://jamanetwork.com/journals/jama/fullarticle/2776543

42. McCarthy KR, Rennick LJ, Nambulli S, Robinson-McCarthy LR, Bain WG, Haidar G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021 Mar 12;371(6534):1139–42.

43. Ip JD, Kok K-H, Chan W-M, Chu AW-H, Wu W-L, Yip CC-Y, et al. Intra-host non-synonymous diversity at a neutralizing antibody epitope of SARS-CoV-2 spike protein N-terminal domain. Clin Microbiol Infect. 2020 Nov;S1198743X20306613.

44. Bangaru S, Ozorowski G, Turner HL, Antanasijevic A, Huang D, Wang X, et al. Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. Science. 2020 Nov 27;370(6520):1089–94.

45. Hirotsu Y, Omata M. Household transmission of SARS-CoV-2 R.1 lineage with spike E484K mutation in Japan [Internet]. Infectious Diseases (except HIV/AIDS); 2021 Mar [cited 2021 May 8]. Available from: http://medrxiv.org/lookup/doi/10.1101/2021.03.16.21253248

46. Cavanaugh AM, Fortier S, Lewis P, Arora V, Johnson M, George K, et al. COVID-19 Outbreak Associated with a SARS-CoV-2 R.1 Lineage Variant in a Skilled Nursing Facility After Vaccination Program — Kentucky, March 2021. MMWR Morb Mortal Wkly Rep. 2021 Apr 30;70(17):639–43.

Figure 1

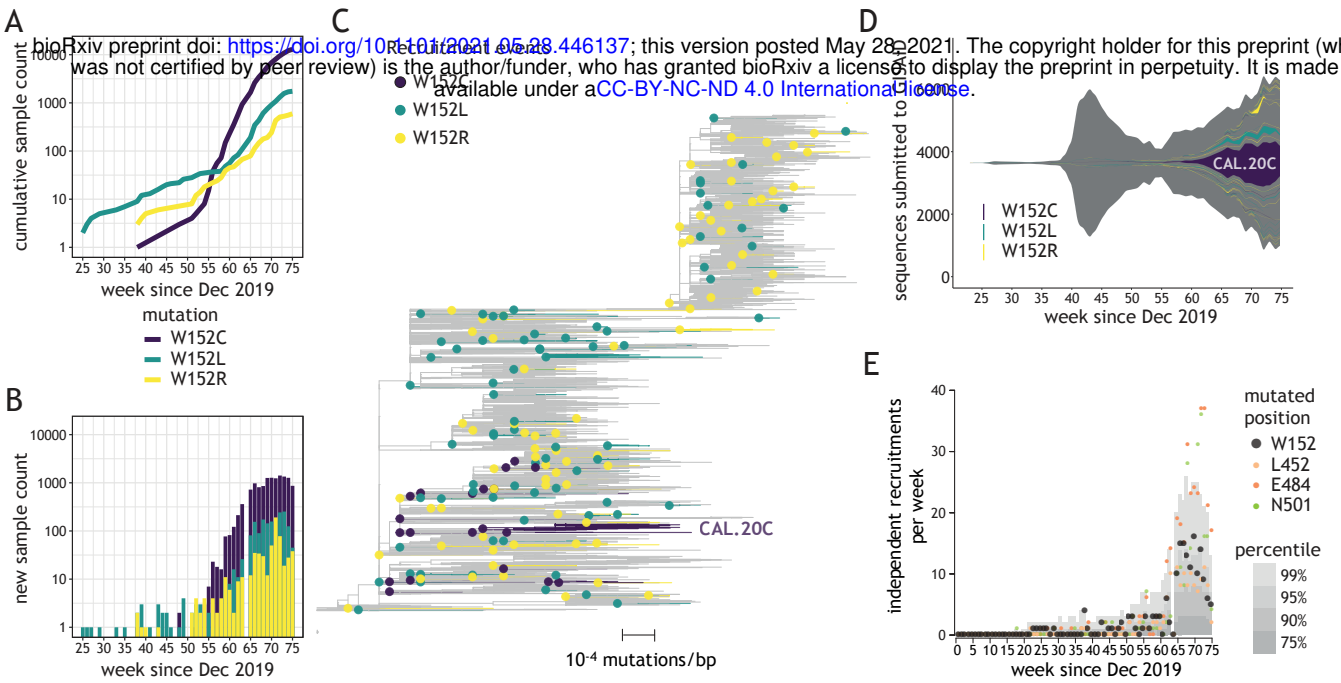
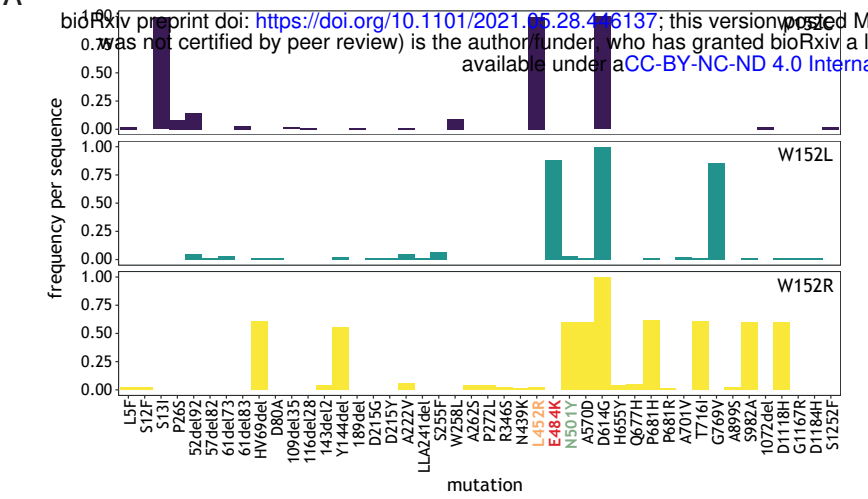
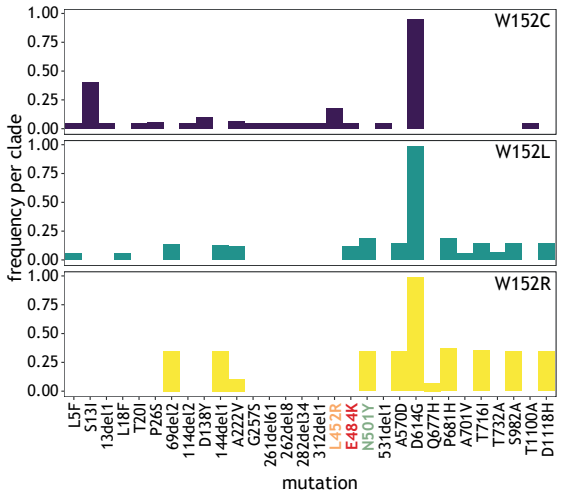


Figure 2





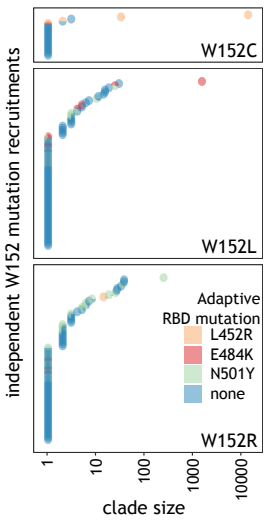


Figure 3

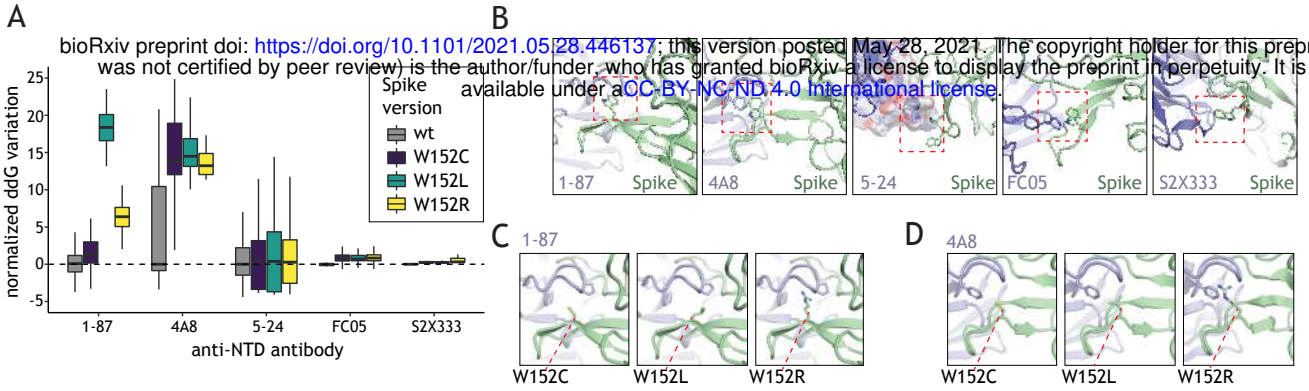


Figure 4