

 Open access • Posted Content • DOI:10.1101/2021.02.22.432189

## Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of SARS-Cov-2 variants — [Source link](#)

Veronika Tchesnokova, Hemantha Kulakesara, Lydia Larson, Victoria Bowers ...+13 more authors

**Institutions:** University of Washington, Kaiser Permanente

**Published on:** 11 Mar 2021 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** Mutation (genetic algorithm) and Lineage (genetic)

Related papers:

- [Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus.](#)
- [Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 \(SARS-CoV-2\) lineage with multiple spike mutations in South Africa](#)
- [The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity.](#)
- [Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor.](#)
- [Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/acquisition-of-the-l452r-mutation-in-the-ace2-binding-40qcl2knec>

1 **Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein**  
2 **triggers recent massive expansion of SARS-Cov-2 variants**

3 Veronika Tchesnokova<sup>1,2</sup>, Hemantha Kulakesara<sup>3</sup>, Lydia Larson<sup>1</sup>, Victoria Bowers<sup>4</sup>,  
4 Elena Rechkina<sup>2#</sup>, Dagmara Kisiela<sup>1\*</sup>, Yulia Sledneva<sup>2</sup>, Debarati Choudhury<sup>2</sup>, Iryna  
5 Maslova<sup>2</sup>, Kai Deng<sup>3</sup>, Kirithi Kutumbaka<sup>3</sup>, Hao Geng<sup>3</sup>, Curtis Fowler<sup>3</sup>, Dina Greene<sup>5</sup>,  
6 James Ralston<sup>5</sup>, Mansour Samadpour<sup>3</sup>, Evgeni Sokurenko<sup>1\$</sup>

7 1 University of Washington, Seattle, WA

8 2 ID Genomics, Inc., Seattle, WA

9 3 IEH Laboratories and Consulting Group, Seattle, WA

10 4 ARMADA (The Antibiotic Resistance Monitoring, Analysis and Diagnostics Alliance),  
11 Seattle, WA

12 5 Kaiser Permanente Washington (KPWA) and KPWA Research Institute, Seattle, WA

13 \$ Corresponding author: Evgeni V. Sokurenko, [evs@uw.edu](mailto:evs@uw.edu)

14

15 **Abstract.**

16 The recent rise in mutational variants of SARS-CoV-2, especially with changes in the  
17 Spike protein, is of significant concern due to the potential ability for these mutations to  
18 increase viral infectivity, virulence and/or ability to escape protective antibodies. Here,  
19 we investigated genetic variations in a 414-583 amino acid region of the Spike protein,  
20 partially encompassing the ACE2 receptor-binding domain (RBD), across a subset of  
21 570 nasopharyngeal samples isolated between April 2020 and February 2021, from  
22 Washington, California, Arizona, Colorado, Minnesota and Illinois. We found that  
23 samples isolated since November have an increased number of amino acid mutations in  
24 the region, with L452R being the dominant mutation. This mutation is associated with a  
25 recently discovered CAL.20C viral variant from clade 20C, lineage B.1.429, that since  
26 November-December 2020 is associated with multiple outbreaks and is undergoing  
27 massive expansion across California. In some samples, however, we found a distinct  
28 L452R-carrying variant of the virus that, upon detailed analysis of the GISAID database  
29 genomes, is also circulating primarily in California, but emerged even more recently.  
30 The newly identified variant derives from the clade 20A (lineage B.1.232) and is named  
31 CAL.20A. We also found that the SARS-CoV-2 strain that caused the only recorded  
32 case of infection in an ape - gorillas in the San Diego Zoo, reported in January 2021 - is  
33 CAL.20A. In contrast to CAL.20C that carries two additional to L452R mutations in the  
34 Spike protein, L452R is the only mutation found in CAL.20A. According to the  
35 phylogenetic analysis, however, emergence of CAL.20C was also specifically triggered  
36 by acquisition of the L452R mutation. Further analysis of GISAID-deposited genomes  
37 revealed that several independent L452R-carrying lineages have recently emerged  
38 across the globe, with over 90% of the isolates reported between December 2020 -  
39 February 2021. Taken together, these results indicate that the L452R mutation alone is  
40 of significant adaptive value to SARS-CoV-2 and, apparently, the positive selection for  
41 this mutation became particularly strong only recently, possibly reflecting viral  
42 adaptation to the containment measures or increasing population immunity. While the  
43 functional impact of L452R has not yet been extensively evaluated, leucine-452 is  
44 positioned in the receptor-binding motif of RBD, in the interface of direct contact with the  
45 ACE2 receptor. Its replacement with arginine is predicted to result in both a much

46 stronger binding to the receptor and escape from neutralizing antibodies. If true, this in  
47 turn might lead to significantly increased infectivity of the L452R variants, warranting  
48 their close surveillance and in-depth functional studies.

## 49 **Introduction**

50 The recent emergence of mutational variants of SARS-CoV-2 (nCoV) around the globe  
51 suggests adaptive evolution of the virus, potentially affecting its transmissibility,  
52 infectivity, virulence and/or immune escape [1-4]. The primary target of current vaccines  
53 and monoclonal antibodies is the Spike protein which mediates viral attachment to and  
54 entry into host cells [5, 6]. Thus, emergence of variants with mutations in the Spike  
55 protein are of particular interest due to their potential for reduced susceptibility to  
56 neutralizing antibodies elicited by vaccination or prior infection.

57 The Spike protein is 1,273 amino acids long and is comprised of the N-terminal region  
58 S1 (amino acids 14-682) responsible for viral attachment to target cells via the ACE2  
59 receptor, and the C-terminal region S2 (aa 686–1273) responsible for membrane fusion  
60 and cell entry [7]. Before fusion, S1 is cleaved from S2 in the cleavage region (aa 682-  
61 685). Antibodies against the ACE2 binding domain of S1 (Receptor-Binding Domain,  
62 RBD; aa 319-541) are considered to be critical in neutralizing nCoV [8-10]. Because of  
63 the important functional and antigenic properties of RBD, structural changes in this  
64 domain deserve special attention and have already been highlighted by such notorious  
65 RBD mutations as E484K (e.g. found in the ‘Brazil’ variant B.1.1.28) or N501Y (found in  
66 the ‘British’ B.1.1.7 variant and ‘South African’ B.1.351 variant) [1].

67 We evaluated the feasibility of determining mutational changes in the S1 region by PCR  
68 amplification of 541 bases fragment within and immediately downstream from the RBD  
69 coding region, followed by Sanger sequencing (see Methods). The amplicon included  
70 the gene region coding aa 414 to 583 of the Spike protein (for further: ‘region 414-583’),  
71 which includes the so-called receptor-binding ridge epitope (aa 417, 455, 456 and 470–  
72 490), the 443–450 loop epitope (aa 443–452 and 494–501) and the 570-572 loop of the  
73 so-called C-terminal domain 1 (CTD1) of S1, which is important in the interaction  
74 between the S1 and S2 regions. As test samples, we used 570 clinical oropharyngeal

75 specimens collected during April-May from nCoV-positive patients at Kaiser  
76 Permanente Washington (51 samples) and nCoV-positive clinical samples submitted for  
77 testing to IEH Laboratories, Inc. (Bothell, WA), from four states (California, Washington,  
78 Arizona, Colorado, Minnesota and Illinois) that were split into a September-October  
79 collection group (85 samples) and a November-February collection group (434  
80 samples). In the process of region 414-583 analysis, we noted a high prevalence of  
81 samples that carried nCoV variants with L452R mutation. This prompted us to perform  
82 an in-depth follow-up analysis of the L452R-carrying nCoV strains in our samples and,  
83 then, their prevalence and clonal origin on the global scale by using publicly available  
84 nCoV genomes and analytical tools of GISAID and Nextstrain databases.

## 85 **Results**

86 It was possible to amplify the 414-583 region and obtain high quality sequence from  
87 99.8% of the specimens. A total of 58 of the samples had 1 to 2 changes from the  
88 corresponding region of the reference Wuhan-Hu-1 genome (NC\_045512). All  
89 sequence changes in the region were single nucleotide polymorphisms (SNPs). Among  
90 the April-May samples, only one sample (2.0%) was found with a mutation. This  
91 mutation was of silent (synonymous) nature, i. e. without the change in amino acid  
92 content (**Table 1**). Among the September-October samples, four samples (4.7%) had  
93 single SNPs, including three silent mutations and one amino acid (nonsynonymous)  
94 mutation. Among the November-February samples, 53 (12.2%) had a total of 15 silent  
95 SNPs and 11 amino acid (nonsynonymous) mutations. Among the later samples, all  
96 multiple samples tested at the IEH Laboratories that were known to have been isolated  
97 from the same patient or submitted from the same collection site on the same day had  
98 the same mutational profiles and were considered as epidemiologically linked and  
99 considered as duplicated. Upon removal of the duplicates, 39 non-linked samples  
100 contained mutations in the 414-583 region, with 27 unique SNPs or SNP combinations.

101 Silent mutations were distributed across the 414-583 region, without clear clustering  
102 (**Figure 1**). In contrast, the amino acid changes were distributed non-randomly and,  
103 except for one mutation (P463S), were clustered within the main epitope regions of

104 RBD (L452R, T470N, T478K, G482V, E484K and S494P) or in the 570-572 loop of  
105 CTD1 (A570V and T572I). All of the amino acid mutations had been reported previously  
106 and could be identified in nCoV sequences deposited to GISAID. Though the notorious  
107 N501Y mutation was not identified in our samples, five samples contained mutation  
108 E484K that is found in lineages from various countries and present in the 'Brazil' variant  
109 B.1.1.28. E484K is located in the receptor-binding ridge epitope and was shown to  
110 provide marked resistance to neutralizing antibodies in multiple studies [8, 11, 12].

111 The most frequent mutation found in the region 414-583 was by far L452R, occurring in  
112 19 out of 39 (48.7%) total and 13 out of 26 (50.0%) non-linked samples, isolated in 9 out  
113 of 18 (50%) separate collection sites. Three samples with L452R also had separate  
114 silent changes and two samples (from the same site) had an additional amino acid  
115 change T572I. L452R was found mostly in samples (14) and sites (6) from California,  
116 though 5 samples from 3 sites came from Washington. L452R is part of the 443-450  
117 loop epitope and is located on the edge of the receptor-binding motif of RBD formed by  
118 residues in direct contact with ACE2 [8, 13, 14]. Its occurrence appeared to be only  
119 sporadic in most of 2020 and has received significant attention only very recently due to  
120 the report of the California Department of Public Health released January 17, 2021 and  
121 a follow-up publication [15]. It was noted that there has been a recent sharp rise in  
122 isolation of nCoV variants with L452R across multiple outbreaks in California,  
123 accounting for more than a third of all isolates. According to the GISAID database, only  
124 six nCoV genomes with L452R were deposited in September-October 2020 (all from  
125 California), but since then additional genomes with the mutation have been deposited  
126 including 142 in November (95.7% from California), 488 in December (79.1%) and 619  
127 in January 2021 (69.2%). This expansion has been linked to a single viral variant from  
128 clade 20C according to the Nextstrain nomenclature of nCoV (lineage B.1.429  
129 according to the PANGO nomenclature of nCoV) and was designated as CAL.20C  
130 (20C/S:452R; B.1.429).

131 According to genomic analysis, CAL.20C has also been defined by 4 additional amino  
132 acid mutations, including two Spike protein mutations, S13I and W152C, located in the  
133 signal peptide and N-terminal domain, respectively [15]. Using an approach similar to

134 our sequencing of the region 414-583, we amplified and sequenced the aa1-250 coding  
135 region in all samples with the L452R mutation. Both S13I and W152C mutations were  
136 found in 15 total (11 non-linked) samples suggesting their identity with CAL.20C.  
137 (Among those samples, one contained an additional silent mutation, while in two  
138 samples from one site a four-amino acid deletion (141-144 LGVY) was found.)  
139 Surprisingly, in 4 samples with L452R the additional mutations were absent. Those  
140 samples originated from two separate sites (two samples in each) in California. One  
141 sample pair carried the T572I mutation in the 414-583 region mentioned above. To  
142 determine how closely these S13I/W152C non-carrier L452R variants were related to  
143 the CAL.20C variant, full genome sequencing was possible performed on three of those  
144 samples and on four CAL.20C-like, S13I/W152C carrier samples. Based on the  
145 genome-wide analysis, all CAL.20C-like variants were in the same clade as the chosen  
146 reference CAL.20C strain (GISAID # 730092) isolated in September, 2020 (**Figure 2A**).  
147 In sharp contrast, the other L452R variants formed a distinct phylogenetic clade that is  
148 distant from CAL.20C, sharing none of the CAL.20C-specific mutations. In fact, further  
149 analysis established that those strains derived from a separate 20A clade, lineage  
150 B.1.232, indicating that the L452R mutation in this group of strains was acquired  
151 independently from CAL.20C. To distinguish from the recently described CAL.20C  
152 variant, we designated this novel L452R-carrying variant as CAL.20A  
153 (20A/S:452R/B.1.232).

154 Analysis of the GISAID database on February 19, 2021, revealed that the CAL.20A-  
155 including lineage B.1.232 contained a total of 559 deposited genomes, but only 54 of  
156 them (9.7%) contained the L452R mutation, all closely related to the CAL.20A strains  
157 identified here. The first deposition of an CAL.20A genome was made on November 23,  
158 2020, from Baja California, Mexico (GISAID # 878300), with most of the CAL.20A  
159 genomes deposited in January-February 2021 (43 genomes; 79.6%). The vast majority  
160 of CAL.20A strains (74%) were isolated in the state of California, but also in 5 other  
161 states (WI, NM, UT, PA, AZ) as well as Canada, Mexico and Costa-Rica. Interestingly,  
162 one of the CAL.20A samples was derived from a gorilla in the San Diego Zoo  
163 (deposited to GISAID January 10; GISAID # 862722) – a highly publicized case of nCoV  
164 infection in apes [16].

165 To compare the clonal diversity of CAL.20A and CAL.20C strains, we build a  
166 phylogenetic tree of the January isolates of CAL.20A and 50 randomly selected  
167 genomes of CAL.20C also deposited in January 2021. We also included into the  
168 analysis some strains that we have identified as the most closely related to the L452R  
169 variants, but without this mutation. Based on the shorter branches overall, the CAL.20A  
170 cluster appeared to be less diverse than the CAL.20C cluster (**Figure 2B**). The pairwise  
171 difference in the number of silent, presumably neutrally accumulating mutations per  
172 genome was  $2.56 \pm 1.41$  and  $8.22 \pm 4.19$  mutations, respectively ( $P < .01$ ), indicating that  
173 CAL.20A has emerged much more recently than CAL.20C.

174 In contrast to CAL.20C, L452R was the only omnipresent amino acid mutation in the  
175 Spike protein of CAL.20A, relatively to the Wuhan-Hu-1 reference strain. Other  
176 mutations, like T572I in our samples, were found only sporadically and in few strains  
177 (with most mutations in the S2 region). In fact, L452R was the only amino acid mutation  
178 in the entire genome that separated CAL.20A from the closest non-L452R strains within  
179 the B.1.232 lineage (GISAID # 636127, **Figure 2B**). Thus, acquisition of L452R appears  
180 to be the primary evolutionary event that led to emergence of CAL.20A. Though it was  
181 originally reported that, besides L452R, all CAL.20C strains carry 4 more specific amino  
182 acid mutations, including S13I and W152C in the Spike protein, we found that few  
183 closely related strains in the B.1.429 lineage carry either S13I alone (GISAID # 977963)  
184 or both S13I and W152 but not L542R (GISAID # 847642 and # 977918, **Figure 2B**).  
185 Thus, the Spike mutations in the CAL.20C-containing lineage were acquired sequentially  
186 with the L452R acquired most recently, triggering the current massive clonal expansion  
187 of CAL.20C.

188 Examination of the Nextstrain and GISAID databases on February 18, 2021, revealed  
189 that, besides CAL.20A and CAL.20C, there are a total of 410 nCoV genomes that  
190 contain the L452R mutation which has been acquired by at least 5 separate lineages –  
191 A.21, A.2.4, B.1.1.10, B.1.1.130 and C.16 within different clades, ranging from 2 to 213  
192 strains each (**Figure 3**). The strains were isolated from over 20 countries across all  
193 continents, with no apparent dominance of any geographic area. In one deposited  
194 strain, from lineage B.1.74, an L452Q mutation was present instead of L452R. There is



195 a clear temporal trend in the number of deposited genomes, with 2 genomes only from  
196 July-September, 28 from October-November, 127 from December 2020 and 238 from  
197 January 2021. Thus, the L452R mutation has been acquired independently in a variety  
198 of clonally diverse nCoV strains, with 92.6% of L452R strains reported after November  
199 2020, indicating a very recent emergence of all those lineages.

## 200 **Discussion**

201 Taken together, our results show that two independent nCoV variants recently emerged  
202 in the state of California that carry the L452R mutation in the Spike protein, the already  
203 defined and currently dominant CAL.20C [15] as well as the more recently emerged  
204 CAL.20A identified here. The fact that, according to our analysis, emergence of both  
205 CAL.20A and CAL.20C was triggered by the L452R mutation alone provides direct  
206 evidence for the adaptive significance of this mutation specifically and, also, creates a  
207 potential opportunity to isolate and functionally compare naturally occurring isogenic  
208 variants of nCoV with and without L452R.

209 On the other hand, it is possible that the lack of other mutations specific to CAL.20A or  
210 shared with CAL.20C could be the reason why the former variant has not undergone as  
211 extensive expansion as CAL.20C. This would indicate (and availability of CAL.20A  
212 should help to affirm) that other mutations in CAL.20C might enhance the adaptive  
213 impact of L452R, i.e. that the genomic background of L542R plays a significant role as  
214 the target of positive selection. However, it is also possible that CAL.20A is at least as fit  
215 as CAL.20C, but has not expanded as broadly because it emerged much more recently  
216 and after CAL.20C has underwent extensive expansion in the same geographic niche  
217 area. Co-circulation of CAL.20A and CAL.20C in the same area provides a unique  
218 opportunity to study the interplay between variants in space and time in  
219 demographically diverse but interconnected communities and patient populations.

220 According to the public databases, in addition to the two California strains the L452R  
221 mutation has been acquired at this point by at least half a dozen independent lineages  
222 across multiple countries and continents. Though detailed examination of the timing,  
223 geography and genomic background of L452R emergence in different lineages is

224 beyond the scope of this study, such repeatedly emerging hot-spot mutations typically  
225 indicate strong positive selection. Interestingly, it appears that the selection for L452R  
226 became especially strong very recently. This is possibly reflecting adaptive evolution of  
227 the virus in response to either the epidemiological containment measures extensively  
228 introduced in the fall of 2020 or a growing proportion of the population with immunity to  
229 the original viral variants, i.e. the convalescents and vaccinated individuals.

230 Though potentially an accidental event, isolation of CAL.20A from a gorilla at the San  
231 Diego Zoo is worthy of note. According to the sequence deposited in GISAID, the gorilla  
232 CAL.20A variant carries two additional SNPs, both in the ORF1ab non-structural protein  
233 2 (nsp2), a silent c934t and non-synonymous c810t (T183I), but with several sequence  
234 stretches unfortunately missing. It is impossible to say at this point to what extent the  
235 isolation of the CAL.20A variant is connected to possibly distinctive biological properties  
236 of the strain and specifically the L452R mutation. However, it would be valuable to  
237 determine whether the occurrence of CAL.20A infection in the gorilla is due to specific  
238 features of CAL.20A with regard to viral transmissibility, infectivity or virulence.

239 Despite mounting evidence for the adaptive value of L452R, the exact functional or  
240 structural effect of the mutation and its impact on viral immunogenicity, pathogenesis,  
241 infectivity and/or transmissibility remains to be determined. Due to only recent attention  
242 to L452R variants, relatively few studies investigated the potential effects of this  
243 mutation. It was found that L452R reduces the Spike protein reactivity with a panel of  
244 the virus neutralizing antibodies and sera from convalescent patients [8, 14]. Moreover,  
245 it was found that out of 52 naturally occurring mutations in the receptor-binding motif  
246 residues of RBD that form the interface of direct interaction with ACE2, L452R results in  
247 the largest increase in free energy of the RBD-ACE2 binding complex, predicting  
248 stronger virus-cell attachment and, thus, increased infectivity [13]. We hope that the  
249 current and the original report on L452R variants of nCoV will facilitate in-depth  
250 structure-functional studies of the leucine residue in position 452 position and the  
251 change of hydrophobic leucine to a polar, highly hydrophilic arginine (or glutamine).

252 We cannot exclude the possibility of sample collection bias in our study as it was not  
253 originally designed as an in-depth surveillance study in specific geographical regions. In  
254 addition, our analysis is limited to a relatively small set of samples in hand and to  
255 publicly available genomes. However, we believe that the identification of CAL.20A and  
256 CAL.20C, with both common and unique features relative to the other circulating nCoV  
257 variant, will be useful in the optimization of real-time monitoring and the more complete  
258 understanding of the biological properties of this pandemic virus with a recently  
259 expanding number of genetic variations that are cause for significant public concern.

## 260 **Acknowledgement**

261 We thank Prof. Steven Moseley for critical proofreading of the manuscript and scientific  
262 advice; clinical laboratory staff at Kaiser Permanente Washington and IEH Laboratories  
263 and Consulting Group for providing their support and technical expertise.

## 264 **Funding**

265 The study was funded by NIH Grant R01AI106007, ARMADA foundation, recapture  
266 funds of Prof. E. Sokurenko laboratory and corporate funds of ID Genomics Inc. and  
267 IEH and Consulting Group.

## 268 **Reference**

269 1. Mascola JR, Graham BS, Fauci AS. SARS-CoV-2 Viral Variants-Tackling a Moving  
270 Target. JAMA. 2021 Feb 11. doi: 10.1001/jama.2021.2088. Epub ahead of print. PMID:  
271 33571363.

272 2. Flores-Alanis A, Cruz-Rangel A, Rodríguez-Gómez F, González J, Torres-Guerrero  
273 CA, Delgado G, Cravioto A, Morales-Espinosa R. Molecular Epidemiology Surveillance  
274 of SARS-CoV-2: Mutations and Genetic Diversity One Year after Emerging. Pathogens.  
275 2021 Feb 9;10(2):184. doi: 10.3390/pathogens10020184. PMID: 33572190.

276 3. Luring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2-What Do They Mean?  
277 JAMA. 2021 Feb 9;325(6):529-531. doi: 10.1001/jama.2020.27124. PMID: 33404586.

- 278 4. Graudenzi A, Maspero D, Angaroni F, Piazza R, Ramazzotti D. Mutational signatures  
279 and heterogeneous host response revealed via large-scale characterization of SARS-  
280 CoV-2 genomic diversity. *iScience*. 2021 Feb 19;24(2):102116. doi:  
281 10.1016/j.isci.2021.102116. Epub 2021 Jan 28. PMID: 33532709; PMCID:  
282 PMC7842190.
- 283 5. Izda V, Jeffries MA, Sawalha AH. COVID-19: A review of therapeutic strategies and  
284 vaccine candidates. *Clin Immunol*. 2021 Jan;222:108634. doi:  
285 10.1016/j.clim.2020.108634. Epub 2020 Nov 17. PMID: 33217545; PMCID:  
286 PMC7670907.
- 287 6. Sharma O, Sultan AA, Ding H, Triggler CR. A Review of the Progress and Challenges  
288 of Developing a Vaccine for COVID-19. *Front Immunol*. 2020 Oct 14;11:585354. doi:  
289 10.3389/fimmu.2020.585354. PMID: 33163000; PMCID: PMC7591699.
- 290 7. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veersler D. Structure, Function,  
291 and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*. 2020 Apr 16;181(2):281-  
292 292.e6. doi: 10.1016/j.cell.2020.02.058. Epub 2020 Mar 9. Erratum in: *Cell*. 2020 Dec  
293 10;183(6):1735. PMID: 32155444; PMCID: PMC7102599.
- 294 8. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD.  
295 Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that  
296 affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe*. 2021 Feb  
297 8:S1931-3128(21)00082-2. doi: 10.1016/j.chom.2021.02.003. Epub ahead of print.  
298 PMID: 33592168; PMCID: PMC7869748.
- 299 9. Lu S, Xie XX, Zhao L, Wang B, Zhu J, Yang TR, Yang GW, Ji M, Lv CP, Xue J, Dai  
300 EH, Fu XM, Liu DQ, Zhang L, Hou SJ, Yu XL, Wang YL, Gao HX, Shi XH, Ke CW, Ke  
301 BX, Jiang CG, Liu RT. The immunodominant and neutralization linear epitopes for  
302 SARS-CoV-2. *Cell Rep*. 2021 Jan 26;34(4):108666. doi: 10.1016/j.celrep.2020.108666.  
303 PMID: 33503420; PMCID: PMC7837128.
- 304 10. Ku Z, Xie X, Davidson E, Ye X, Su H, Menachery VD, Li Y, Yuan Z, Zhang X,  
305 Muruato AE, I Escuer AG, Tyrell B, Doolan K, Doranz BJ, Wrapp D, Bates PF, McLellan

- 306 JS, Weiss SR, Zhang N, Shi PY, An Z. Molecular determinants and mechanism for  
307 antibody cocktail preventing SARS-CoV-2 escape. *Nat Commun.* 2021 Jan  
308 20;12(1):469. doi: 10.1038/s41467-020-20789-7. PMID: 33473140; PMCID:  
309 PMC7817669.
- 310 11. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler  
311 M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. Neutralization of  
312 SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-  
313 elicited sera. *Nat Med.* 2021 Feb 8. doi: 10.1038/s41591-021-01270-4. Epub ahead of  
314 print. PMID: 33558724.
- 315 12. Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico  
316 JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH,  
317 Diamond MS, Whelan SPJ. Identification of SARS-CoV-2 spike mutations that attenuate  
318 monoclonal and serum antibody neutralization. *Cell Host Microbe.* 2021 Jan 27:S1931-  
319 3128(21)00044-5. doi: 10.1016/j.chom.2021.01.014. Epub ahead of print. PMID:  
320 33535027; PMCID: PMC7839837.
- 321 13. Chen J, Wang R, Wang M, Wei GW. Mutations Strengthened SARS-CoV-2  
322 Infectivity. *J Mol Biol.* 2020 Sep 4;432(19):5212-5226. doi: 10.1016/j.jmb.2020.07.009.  
323 Epub 2020 Jul 23. PMID: 32710986; PMCID: PMC7375973.
- 324 14. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H,  
325 Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. The Impact of  
326 Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell.* 2020 Sep  
327 3;182(5):1284-1294.e9. doi: 10.1016/j.cell.2020.07.012. Epub 2020 Jul 17. PMID:  
328 32730807; PMCID: PMC7366990.
- 329 15. Zhang W, Davis BD, Chen SS, Sincuir Martinez JM, Plummer JT, Vail E.  
330 Emergence of a Novel SARS-CoV-2 Variant in Southern California. *JAMA.* 2021 Feb  
331 11. doi: 10.1001/jama.2021.1612. Epub ahead of print. PMID: 33571356.

332 16. Gibbons A. Captive gorillas test positive for coronavirus. *Science*, January 12,  
333 2021. doi:10.1126/science.abg5458.  
334 <https://www.sciencemag.org/news/2021/01/captive-gorillas-test-positive-coronavirus>.

## 335 **METHODS**

### 336 *Sample collection*

337 Random de-identified nasopharyngeal samples were provided either as original swabs  
338 by Kaiser Permanente Washington (KPWA, April - May 2020, greater Seattle area) or  
339 as purified RNA by IEH Laboratories and Consulting group (September 2020 –  
340 February 2021, multiple states). Samples that tested positive for the presence of SARS-  
341 CoV-2 RNA (according to the respective laboratory practices) were subjected to further  
342 analysis. IEH samples of interest (see below) were assigned a unique identifier based  
343 on their collection date and source; information regarding the state of origin was  
344 provided for these samples for epidemiological analysis. Protected health information  
345 was completely removed from all clinical results before they were given to the  
346 researchers. All subjects cannot be identified directly or indirectly. The Western  
347 Institutional Review Board (Puyallup, WA) provided institutional biosafety committee  
348 services to Institute for Environmental Health by approving consent forms and human  
349 research safety protocols.

### 350 *RNA isolation and SARS-Cov-2 testing*

351 RNA from KPWA samples was isolated using both AllPrep DNA/RNA Kit (Qiagen) and  
352 MagMax Viral Pathogen RNA isolation Kit (Thermo Fisher Scientific) according to  
353 manufacturer's procedure. RNA isolation from IEH samples was performed according to  
354 laboratory guidelines using Thermofisher Kingfisher-96 instrument and proprietary IEH  
355 Nucleic Acid Extraction Reagent Kit. RNA was stored at -20oC until use. Testing of RNA  
356 for the presence of SARS-Cov-2 RNA at KPWA and IEH was performed according to  
357 laboratory guidelines using CDC 2019 Novel Coronavirus (2019-nCoV) Real-Time  
358 Reverse Transcriptase (RT)–PCR Diagnostic panel and proprietary IEH SARS-CoV-2  
359 RT-PCR Test kit (based on the CDC RT-PCR kit), respectively.

360 *Amplification and sequencing of 414-583 and 1-250 regions*

361 To amplify the 414-583 RBD region of the Spike gene from SARS-CoV-2 RNA samples,  
362 we used both one-step RT-PCR and two-step (cDNA synthesis followed by separate  
363 PCR amplification) reaction designs with commercial One-Step Ahead RT-PCR Kit  
364 (Qiagen) or IEH in-house RT, PCR and RT-PCR kits according to manufacturer's  
365 guidelines. Both kits and conditions yield non-distinguishable results. Primers to amplify  
366 the region were as following: PF, 5'- GTGACATAGTGTAGGCAATGATG-3', PR, 5'-  
367 TGGTGTAATGTCAAGAATCTCAAG-3'. First round of PCR (either RT-PCR on RNA or  
368 PCR on cDNA) consisted of 40 cycles of 10 sec 95oC, 15 sec 57oC, 40 sec 72oC. The  
369 product was then diluted 1:50 times with sterile water, and a second round of PCR was  
370 performed for 15 cycles at the same conditions using T7-tailed nested primers to obtain  
371 a single pure product (PF-T7Pro-nested, 5'-  
372 TAATACGACTCACTATAGGGCAAAGTGGAAAGATTGC-3', PR-T7Term-nested, 5'-  
373 GCTAGTTATTGCTCAGCGGCTCAAGTGTCTGTG-3'). Similarly, 1-250 Spike protein  
374 region was amplified firstly using primers PF2, 5'-  
375 CAGAGTTGTTATTTCTAGTGATGTTC-3' and PR2, 5'-  
376 TGAAGAAGAATCACCAGGAGTC-3', followed by the nested PCR on 1:50 diluted  
377 samples with tailed primers PF-T7Pro, 5'-  
378 TAATACGACTCACTATAGGGCAGAGTTGTTATTTCTAGTGATGTTC-3', and PR-  
379 T7Term-nested2, 5'-GCTAGTTATTGCTCAGCGGGAGTCAAATAACTTC-3'. Sanger  
380 sequencing of the amplified region was performed from both ends by Eton Bioscience,  
381 Inc. Sequences were analyzed using BioEdit 7.2 and MEGA 7 Software.

382 *Whole genome sequencing*

383 WGS was performed by IEH on MiSeq Illumina instrument; each sample was subjected  
384 to two individual rounds of sequencing. Sequences were assembled de novo using  
385 proprietary IEH pipeline or using the PATRICK sequence assembly service  
386 (<https://patricbrc.org/app/Assembly2>).

387 *Phylogenetic analysis of SARS-CoV-2 genomes*

388 The latest global analysis of SARS-CoV-2 genomes from the Nextstrain database  
389 (<https://nextstrain.org/ncov/global>) was used to determine viral strains with the presence  
390 of L452 substitutions in the Spike protein (as of 12/17/2021). All genome sequences  
391 from PANGO lineages A.2.4, A.21, B.1.1.10, B.1.1.74, B.1.1.130, B.1.232, B.1.429,  
392 C.16 were downloaded from GISAID database (<https://www.epicov.org/epi3>, as of  
393 12/17/2021). Sequences were checked for the presence of Spike-L452 substitution(s).  
394 All L452R-containing sequences submitted in January-February 2021 for lineages  
395 B.1.232 (CAL.20A) and 50 randomly chosen B.1.429 (CAL.20C) as well as the closely  
396 related L452 ancestors from both lineages were aligned and used to build total and  
397 synonymous-only phylogenetic trees using MEGA 7 software. Deletions of stretches of  
398 nucleotides which resulted in in-frame codon deletions were treated as single event.  
399 Unresolved nucleotides were assigned nucleotide value based on the closest relative.  
400 The same L452R sequences were used to run pairwise comparisons for synonymous  
401 substitutions for each CAL.20A and CAL.20C lineage separately and to calculate the  
402 average pairwise differences in silent mutations.

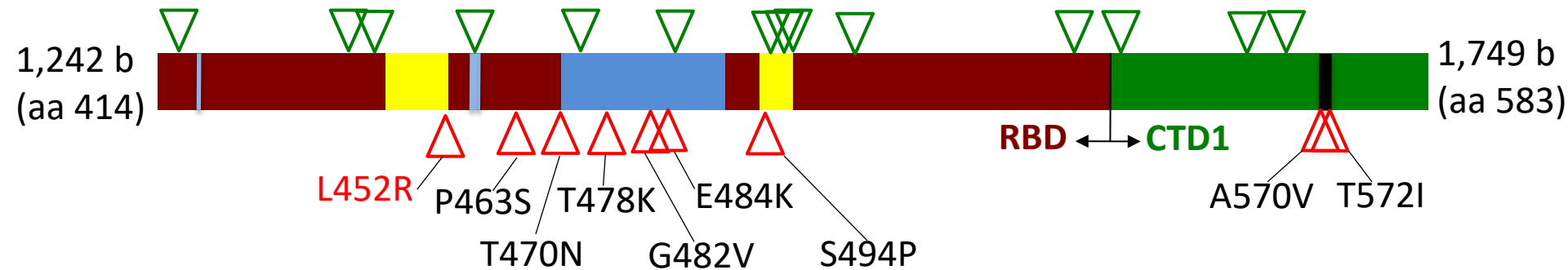
#### 403 *Statistical analysis*

404 Distribution of synonymous (S) and non-synonymous (NS) mutations within the 414-583  
405 fragment was compared for RBD ridge epitope residues (aa 417, 443-452, 455-456,  
406 470-490, 497—501, 570-572, total N = 49) vs other regions (N = 169) in McNemar test.  
407 There were 6 vs 1 NS and 7 and 8 S mutations, respectively, resulting in McNemar Chi2  
408 4.50 (P = .034).



**Table 1. Mutations distribution across the test samples.** No. samples column shows total number of samples from a site received same day. In bold – amino acid changes

Time period	No. samples	Site	State	Region 1 (414-583)	Region 2 (1-250)
<b>April - May (N=51)</b>	1	KP	WA	c1425t	
<b>September - October (N=85)</b>	1	22	CA	t1695c	
	1	4	CA	c1497t	
	1	10	CA	c1497t	
	1	23	PA	g1450a ( <b>E484K</b> )	
<b>November - February (N = 434)</b>	1	3	CA	a1290g	
	1	3	CA	t1326c	
	1	20	MN	t1338g	
	1	6	AZ	t1356g	
	1	4	CA	c1425t	
	1	un	un	t1458c	
	1	15	MN	t1506g	
	2	18	NM	c1524t	
	1	13	WA	c1626t	
	1	7	CO	c1770t	
	1	2	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	1	12	WA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	2	4	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	2	14	WA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	1	17	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	1	16	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	2	5	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> ) + del 141-144 LGVY
	1	4	CA	t1355g ( <b>L452R</b> )	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> ) + c76t
	1	4	CA	t1355g ( <b>L452R</b> ) + c1416t	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	2	8	WA	t1355g ( <b>L452R</b> ) + t1593c	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	1	3	CA	t1355g ( <b>L452R</b> ) + c1626t	g38t ( <b>S13I</b> ) + g456t ( <b>W152C</b> )
	2	5	CA	t1355g ( <b>L452R</b> )	
	2	4	CA	t1355g ( <b>L452R</b> ) + t1715g ( <b>T572I</b> )	
	1	21	GA	g1365t ( <b>L455F</b> )	
	1	4	CA	c1387t ( <b>P463S</b> ) + t1503c	
	1	7	CO	c1409a ( <b>T470N</b> )	
	5	4	CA	c1433a ( <b>T478K</b> )	
	1	19	CA	c1433a ( <b>T478K</b> )	
	1	10	CA	c1433a ( <b>T478K</b> ) + c1686t	
	1	1	MN	g1445t ( <b>G482V</b> )	
	1	1	MN	g1450a ( <b>E484K</b> )	
	2	20	MN	g1450a ( <b>E484K</b> )	
	1	13	WA	g1450a ( <b>E484K</b> )	
1	9	WA	t1480c ( <b>S494P</b> )		
1	7	CO	g1696a ( <b>G566S</b> )		
2	9	WA	c1709t ( <b>A570V</b> )		



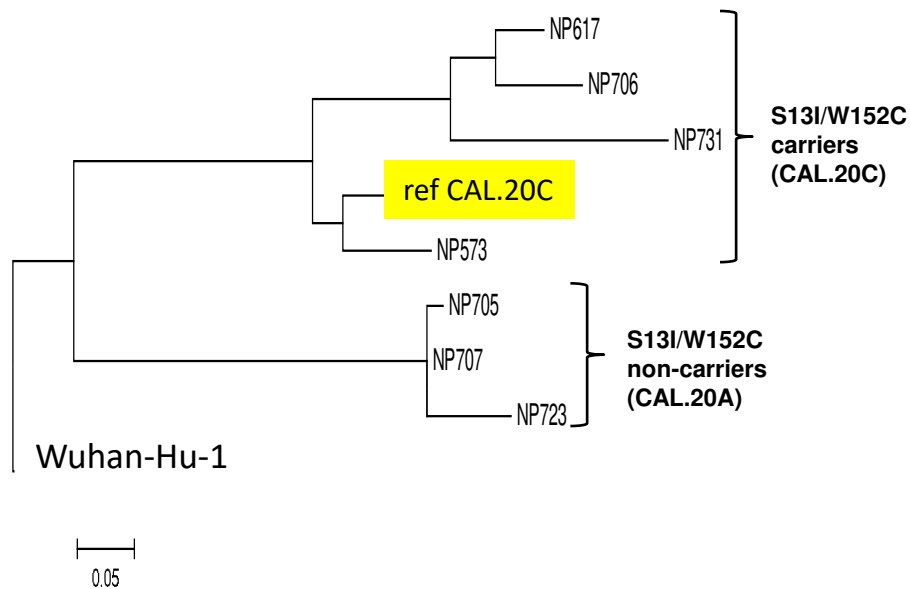
**FIGURE 1: Distribution of silent (green triangles) and amino acid (red triangles) mutations across region 414-583 of the Spike protein.** Dark red – receptor-binding domain (RBD); Green – C-terminal domain 1 (CTD1) of S1 Spike region; Blue – receptor-binding ridge epitope residues; Yellow – 443-450 loop epitope residues; Black – 570-572 loop residues.

**Figure 2. Phylogenetic trees of CAL.20A and CAL.20C genomes.**

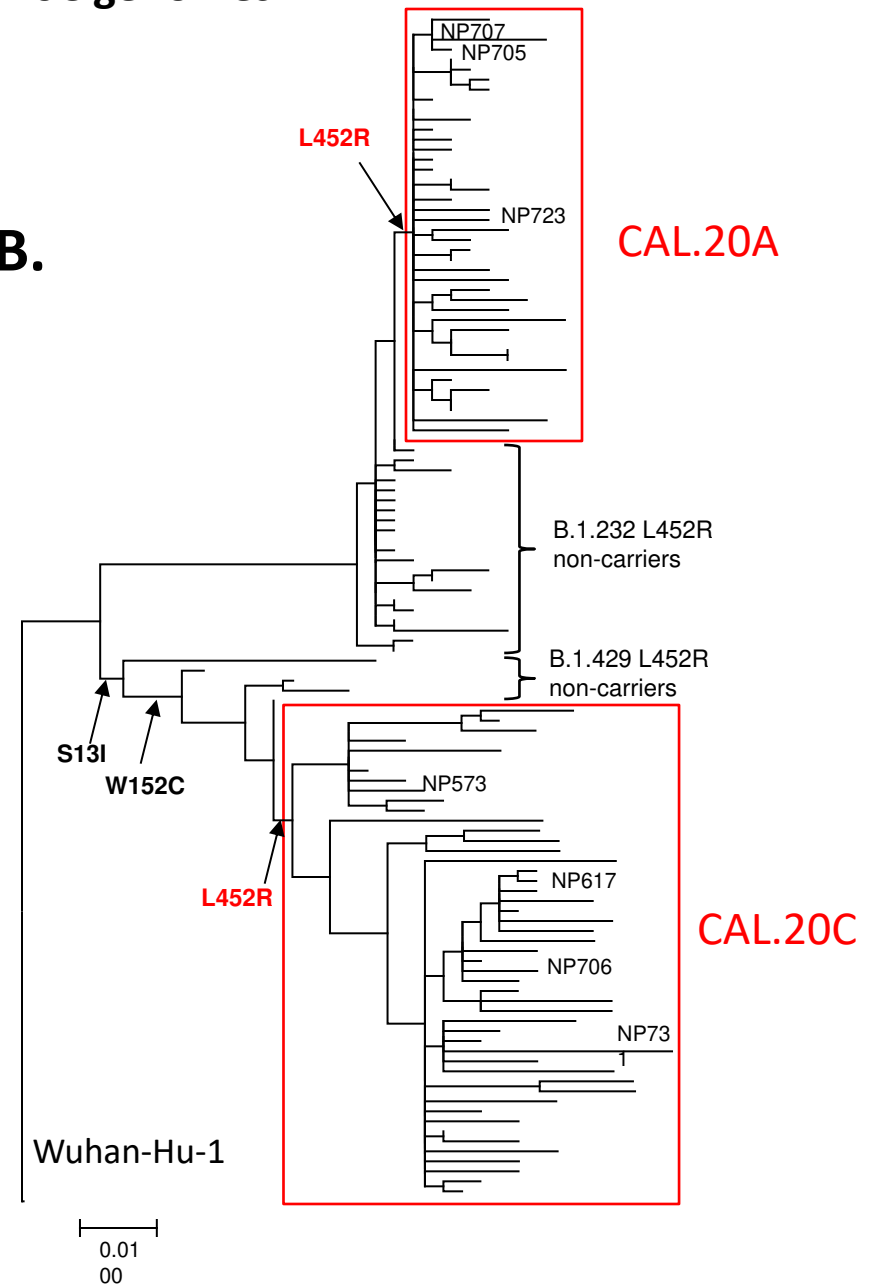
A. nCoV strains identified in the tested samples.

B. nCoV strains deposited into GISAID database.

**A.**



**B.**



**FIGURE 3.**  
**Nextstrain Unrooted**  
**cladogram**  
**phylogenetic tree of**  
**SARS-CoV-2 genetic**  
**variants (as of**  
**February 15 2021).**

Nodes in turquoise -  
L452; in yellow - R452;  
in black - Q452.

Nomenclature:

Nextstrain

clade/PANGO lineage.

At the time of analysis,

CAL.20A strains were

not found in the

Nextstrain database

and only one B.1.232

(non-CAL.20A) was

found.

