Genomic surveillance revealed prevalence of unique SARS-CoV-2 variants bearing mutation in the RdRp gene among Nevada patients

Paul D. Hartley[1,4*], Richard L. Tillett[7*], Yanji Xu[4,5], David P. AuCoin[3,4], Joel R. Sevinsky[6], Andrew Gorzalski[2,4], Mark Pandori[2,4], Erin Buttery[8], Holly Hansen[8], Michael A. Picker[8], Cyprian C. Rossetto[3,4*#], Subhash C. Verma[3,4*#]

[1] Nevada Genomics Center
[2] Nevada State Public Health Laboratory
[3] Department of Microbiology & Immunology, University of Nevada, Reno School of Medicine
[4] University of Nevada, Reno
[5] Nevada Center for Bioinformatics
[6] Theiagen Consulting, LLC
[7] Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas,
[8] Southern Nevada Public Health Laboratory of the Southern Nevada Health District, Las Vegas

* These authors contributed equally to this work and are listed in alphabetical order

[#] To whom correspondence should be addressed:

Cyprian C. Rossetto crossetto@med.unr.edu, Subhash C. Verma scverma@med.unr.edu

Running title: Unique RdRp variant of SARS-CoV-2 in Nevada patients

Keywords: SARS-CoV-2, COVID-19, genome enrichment, nsp12, RdRp, orf1b 314

ABSTRACT:

Background:

SARS-CoV-2 is the etiological agent of coronavirus disease 2019 (COVID-19), which has a wide range of disease manifestations, including relatively mild respiratory illness to severe disease requiring hospitalization. Patients with signs of COVID-19 (fever, coughing, shortness of breath) were tested for the virus by targeting specific SARS-COV-2 genomic loci via RT-PCR using RNA extracted from nasopharyngeal/nasal swabs. However, this testing does not yield genetic information required for identifying viral evolution or variants.

Methods:

We sequenced 200 specimens from positively identified cases in Nevada through our robust protocol for sequencing SARS-CoV-2 genomes directly from patients nasopharyngeal/nasal swabs. This protocol enabled identification of specific nucleotide variants including those for D614G and clade defining mutations. Additionally, these sequences were used to determine the phylogenetic relationships of SARS-CoV-2 genomes of public health importance occurring in the state of Nevada.

Results:

Our study reports the finding of a novel variant in the nsp12 (RdRp-RNA dependent RNA Polymerase) protein at residue 323 (314aa of orf1b) to Phenylalanine (F) from Proline (P), present in the original isolate of SARS-CoV-2 (Wuhan-1). This 323F variant is found at a very high frequency in Northern Nevada suggesting for a founder effect. Structural modeling of the RdRP with P323F mutation did not show any significant difference in

protein conformation but the phenotypic effect is unknown and an area of active investigation.

Conclusions:

Our results highlight the introduction and spread of specific SARS-CoV-2 variants at very high frequency within a distinct geographic location that is important for clinical and public health perspectives in understanding the evolution and transmission of SARS-CoV-2.

BACKGROUND:

Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2), the cause of coronavirus disease 2019 (COVID-19), was first identified and reported in December 2019 in Wuhan, Hubei province, China [1-3]. RNA sequencing and phylogenetic analysis of a specimens taken during the initial outbreak in Wuhan determined that the virus is most closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus Betacoribavirus, subgenus Sarbecovirus) which had previously been identify in bats in China [2, 4]. Coronaviruses have a recent history as emerging infections, first SARS-CoV in 2002-2003, and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012, both zoonotic infections that cause severe respiratory illness in humans [2, 5-8] [2,5-8]. Unlike SARS-CoV and MERS-CoV which displayed limited global spread, SARS-CoV-2 has spread around the world within a few months. There are specific characteristics of SARS-CoV-2 which have facilitated the transmission, including infections that result in asymptomatic or mild disease, allowing for under-characterized transmission.

72 SARS-CoV-2 is an enveloped, positive single-stranded RNA virus. Detection of
73 SARS-CoV-2 in patients has primarily occurred using RT-qPCR to detect viral RNA from
74 respiratory specimens (primarily nasal, nasopharyngeal swabs). While RT-PCR results
75 can be quantified through determination of a cycle threshold (Ct) value for each sample
76 it does not yield sequence data leading to the description of genomic variants. To further
77 a study of such variants, and to better understand the epidemiology of the virus in the
78 state of Nevada, we developed a workflow that allowed us to sequence SARS-CoV-2
79 genomic RNA from patient swabs containing a broad range of viral loads. Of the
80 sequences of SARS-CoV-2 currently submitted to common database, several were
81 obtained after the virus had been passed in Vero cells [9, 10] and others came directly
82 from patient specimens [11, 12]. Certain data has suggested a potential for lab acquired
83 mutations following passage in cell culture [9, 13]. Specifically, a report of SARS-CoV-2
84 passage in Vero cells which resulted in a spontaneous 9 amino acid deletion within the
85 spike (S) protein that overlaps with furin cleavage site [13]. The loss of this site is
86 suggested to increase the viral entry into Vero cells [14]. For both research and
87 epidemiological purposes, sequencing of SARS-CoV-2 directly from patient specimens
88 not only reduces the possibility of lab-adapted mutations following passage in cell culture
89 but also reduces the time that would be spent growing the virus from the patient
90 specimens and subsequently also reduces handling larger amounts of infectious virus.
91 Additionally, one of the goals in developing an optimized SARS-CoV-2 NGS protocol was
92 to be able to generate adequate depth of coverage of the viral genome while minimizing
93 the sequencing of non-viral RNA which would allow for more specimens to be multiplexed
94 together during sequencing.

95 Our workflow employs a combination of RNA amplification, conversion into
96 Illumina-compatible sequencing libraries and enrichment of SARS-CoV-2 library
97 molecules prior to sequencing. Using this novel methodology, we sequenced SARS-
98 CoV-2 from a total of 200 patient specimens collected over a 3-month period originating
99 from Nevada. Of the 200 selected, 173 were sequenced with enough quality to be used
100 for determining SARS-CoV-2 nucleotide variants to perform further phylogenetic analysis
101 and study the viral epidemiology within the state of Nevada. Analysis of the data suggests
102 a specific epidemiological course for the local epidemic within Northern Nevada. This was
103 characterized by an initial observation of variants closely resembling isolates originating
104 directly from China or Europe. Subsequent to government-mandated period of restrictions
105 on business and social activity, we observed that a viral isolate not seen elsewhere in the
106 world emerged within Northern Nevada cases (nucleotide 14,407 and 14,408). This
107 isolate contains an amino acid change in residue P323L/F of the RdRp (nsp12).
108 Furthermore, we found that sampled viral isolates in Southern Nevada, unlike those in
109 Northern Nevada, closely resembled the makeup of the United States in general.
110
111 **METHODS:**
112 **SARS-CoV-2 specimen and library preparation.** Nasal and Nasopharyngeal swab
113 specimens were received at the Nevada State Public Health Lab (NSPHL) or Southern
114 Nevada Public Health Lab (SNPHL) and RNA extraction was completed using either a
115 QIAamp Viral RNA Mini Kit (QIAGEN) or Mag-Bind Viral DNA/RNA kit (Omega Biotek).
116 Specimens were tested for the presence of coronaviral RNA using FDA-approved kits
117 that employed RT-PCR to detect SARS-COV-2 RNA.

118 A set of 200 coronavirus positive specimens were selected for genome
119 sequencing. Specimens were treated with DNase I (QIAGEN) for 30 minutes at room
120 temperature and concentrated using RNeasy Minelute spin columns (QIAGEN) based on
121 the manufacturer supplied protocol. These concentrated samples were converted into
122 Illumina-compatible sequencing libraries with a QIAseq FX Single Cell RNA Library kit
123 (QIAGEN). RNA samples were annealed to a 1:12.5 dilution of QIAseq FastSelect -HMR
124 probes (QIAGEN) to reduce subsequent amplification of human ribosomal RNA. After
125 treatment to remove trace DNA from the samples, a reverse transcription reaction was
126 carried out using random hexamers. The synthesized DNA was ligated to one another,
127 followed by isothermal linear amplification. Amplified DNA (1 μg) was enzymatically
128 sheared to an average insert size of 300 bp, and Illumina-compatible dual-indexed
129 sequencing adapters were ligated to the ends. Next, about 300 ng of adapter-ligated
130 sample was amplified with 6 cycles of PCR with KAPA HiFi HotStart polymerase (Roche
131 Sequencing Solutions). Enrichment of library molecules containing SARS-CoV-2
132 sequence was conducted with a myBaits kit and coronavirus-specific biotinylated probes
133 (Arbor Biosciences). Each enrichment used 500 ng of PCR-amplified DNA, was carried
134 out based on manufacturer instructions at a hybridization temperature of 65° for 16 hours,
135 and was completed with 8-16 cycles of PCR using KAPA HiFi HotStart polymerase.
136 Samples were sequenced using an Illumina Next-seq mid-output (2 x 75). The generated
137 FASTQ files from the sequencing reaction were analyzed as described below. The data
138 files are available at GISAID, NCBI under the
139 https://www.ncbi.nlm.nih.gov/bioproject/657893

140 **Computational analysis.** Sequence pair libraries were trimmed using Trimmomatic,
141 version 0.39 and adapter-clipping setting “2:30:10:2:keepBothReads” [15]. Read pairs
142 were aligned against the Wuhan reference genome (NC_045512.2) by Bowtie 2, version
143 2.3.5, local alignment [16]. PCR optical duplicates were removed via Picard
144 MarkDuplicates [17].

145 Variants were called using Freebayes, version 1.0.2, with ploidy set to 1, minimum
146 allele frequency 0.75, and minimum depth of 4 [18]. No variants were called in the first
147 200 bp and final 63 bp of the COVID-19 genome. High-quality variant sites were selected
148 where site “QUAL > 20” using *vcffilter*, VCFlib version 1.0.0_rc2 [19]. Individual genomes
149 were reconstructed by their filter-passing variants using *bcftools consensus* and only
150 where aligned coverage depth ≥ 4; bases with coverage below four are reported as
151 unknown (Ns) [20].

152 A set of 3,644 complete, high-coverage SARS-CoV-2 genomes reported in the
153 July 15, 2020 Nextstrain.org global subsample and metadata were obtained from GISAID
154 and combined with our own samples to determine their phylogenetic placement [21, 22].
155 Four of the global samples were set aside after screening for unexpected FASTA
156 characters. The combined sets of global and Nevadan samples were aligned together,
157 with metadata, by the *augur* phylodynamic pipelines of the *ncov* build of the *nextstrain*
158 command-line tool, version 2.0.0.post1 [23].

159 Cumulative frequency of D614G, clades (19A, 19B, 20A, 20B, 20C), and P323L/F
160 were calculated at each time point based on the total number of specimens up to the
161 indicated date. Plots and pie charts were generated using GraphPad Prism (version 8).

162 **nsp12 protein modeling.** Sequence of nsp12 (RdRp) protein for SARS-CoV-2
163 (YP_009725307.0) was retrieved from NCBI protein database and 3D model was
164 structured based on a previously published report (PDB ID: 6XEZ [24]. In addition to
165 nsp12 (chain A), the model also contains nsp7 (chain C), nsp8 (chain B and D), nsp13
166 (chain E and F), ligands (ZN, Mg) and RNA template and product strands. Mutational
167 changes to residue 323 within nsp12 were performed using PyMol Molecular Graphics
168 System (version 2.0, Schrödinger LLC). The original proline (P) was mutated to either
169 leucine (L) or phenylalanine (F) as indicated, these residues along with residues
170 containing side chains within 5 Å of P323L/F are shown as sticks. The rotamers for each
171 P323L/F were assessed and those with least rotational strain and steric hindrance were
172 used to generate the final image. To determine any NCBI deposited sequences which
173 contain the P323F variant, standard protein BLAST from the BLASTp suite was used to
174 find nsp12 protein sequences which contained FSTVFP**F**TSFGP (P323F is bold and
175 underline) from full length SARS-CoV-2 genomes. The P323F amino acid changes were
176 confirmed with the NCBI deposited nucleotide sequences.
177
178 **RESULTS:**
179 **RNA-seq workflow and assembly of SARS-CoV-2 genomes.**
180 A total of 200 SARS-CoV-2 positive specimens collected in Nevada from March
181 6th to June 5th were randomly selected to have their viral genomes sequenced for variant
182 analysis and subsequent epidemiological studies (Fig 1A and Materials & Methods). Of
183 the sequenced specimens, 173 had >90% coverage and sufficient depth to accurately
184 call those genomic positions with variants (Fig. 1B). An alignment of SARS-CoV-2

genomes from these specimens is presented as supplemental data (Supplemental Fig. S1) showing 173 specimens with over 90% coverage. These 173 specimens represented 133 patient specimens from Northern Nevada (including Washoe County of which Reno is the major city, the Carson-Tahoe area, and other northern, rural counties), 40 patient specimens from Southern Nevada (Clark county which encompasses Las Vegas). Nucleotide similarity and variants were determined and used to measure the phylogenetic relationships (Supplemental Fig. S2). The combined nucleotide diversity across the entire SARS-CoV-2 genome for the Nevada specimens is shown in figure 1D, along with the genomic areas that were assessed for change in frequency corresponding to amino acids D614G, P323L/F and nucleotide 379.

During the sequencing analysis we also examined the correlation between Ct values from the diagnostic RT-PCR and percentage coverage of the viral genome to determine the performance and robustness of our sequencing method in relation to available viral RNA in a specimen of a given Ct (Fig. 1C). In general, a Ct value less than 31 resulted in at least 90% coverage to the SARS-CoV-2 genome. Importantly, our in-house developed method for viral genome enrichment and sequencing directly from the patient's specimens (nasal and nasopharyngeal swabs) was robust and yielded sequences covering over 90% of the genome even in samples having very high Ct (~40) of viral genome detection. This is highly significant and shows the power of our workflow in sequencing of SARS-CoV-2 genome from a spectrum of samples including the ones having inadequate amounts of specimen (due to the variability in collection) or lower viral loads in nasal secretions. Consequently, our sequencing protocol avoids any molecular epidemiological bias, which may get acquired through cell culture-based amplification

208 especially in those specimens with high Ct (low viral load) as our method eliminates the
209 need of virus culture.

210 **Prevalence of amino acid variant D614G of SARS-CoV-2 spike protein in specimens**
211 **collected in Nevada.** Earlier studies have revealed the emergence, spread and potential
212 importance of an alteration, D614G (genomic change at 23403A>G), of the spike protein
213 [25]. This missense mutation has become a clade-distinguishing locus that differentiates
214 viral isolates originating in Asia from those that have emerged from Europe. A total of 173
215 cases were analyzed to determine the number and relative proportion of the specimens
216 which carried the D614G spike protein variant in Nevada. The cumulative frequency for
217 D614 and G614 were plotted from March 6th to June 5th (Fig. 2A). Specimens from the
218 beginning of March represent the earliest known cases in Nevada, and of the 14
219 specimens sequenced during this time period (March 6-March 15) D614 was the
220 predominant variant. This shifted from March to June with an increasing frequency of the
221 G614 allele. The trend for specimens originating from either Northern Nevada (N-NV) and
222 Southern Nevada (S-NV) both show a higher frequency of G614 (Fig. 2B). We used a
223 subsampling of sequence data from Nextstrain.org to assess the frequency of D614G in
224 the United States and globally during the same time period (March 6th to June 5th) (Fig.
225 2B and 2C). The global trend by continent of D614G is also similar, with G614 at a higher
226 frequency, the one noted exception is in Asia, where D614 and G614 continue to exist in
227 equal proportions (Fig. 2C).

228 **Frequency of SARS-CoV-2 clades in Nevada.** Worldwide, there are currently 5 main
229 clades (19A, 19B, 20A, 20B, 20C) of SARS-CoV-2 differentiated based on specific
230 nucleotide profiles in the Year-letter scheme of https://clades.nextstrain.org. Clade 19A

231 and 19B are defined by C8782T and T28144C respectively. 20A is a derivative of 19A
232 and contains mutations C3037T, C14408T and A23403G (resulting in D614G). 20B is
233 defined by mutations G28881A, G28882A and G28883C, and 20C contains C1059T and
234 G25563T [26].

235 To assess the introduction and spread of the clades in Nevada the cumulative
236 frequency for the clades were plotted from March 6th to June 5th (Fig. 3A). The earliest
237 sequenced specimens from Nevada were collected in the beginning of March (March 6-
238 March 15) and are predominantly from clades 19A and 19B. Additional sequenced
239 specimens collected from March to June revealed a shift to a higher frequency of 20C
240 (Fig. 3A). We performed phylogenetic reconstruction of the Nevada specimens and
241 differentiated the clades on the circular dendrogram by color (Fig. 3B). There were
242 discordant trends in the dominant clade for specimens originating from either Northern
243 Nevada (N-NV) and Southern Nevada (S-NV) (Fig. 3C and 3D). Specimens from
244 Northern Nevada (Washoe County, Carson-Tahoe, and other counties) showed a
245 prevalence of 20C, while the Southern Nevada specimens from Clark county had a larger
246 proportion of 20A (Fig. 3C and Supplemental Fig. S3a). We used a subsampling of
247 Nextstrain.org data to assess the frequency of clades in the United States and globally
248 by continent during the same time period (March 6th to June 5th). The dominant clade in
249 the United States was 20C, similar to the frequency seen in the total Nevada samples (N-
250 NV and S-NV) (Fig. 3A and 3C). The global clade distributions were variable in areas
251 outside of Asia while clades 19A and 19B are noted to be more prevalent in Asia. (Fig.
252 3D).

253 **Prevalence of amino acid variant P323L/F of SARS-CoV-2 nsp12 (RdRp) in Nevada.**
254 Analysis of sequencing data revealed a novel observation for our specimens at bases
255 14,407 and 14,408 which results in a change at residue 323 in nsp12 (RdRp). For the
256 Wuhan isolate at 14,407 and 14,408 there is CC for proline (P), the variants have CT for
257 leucine (P323L) and TT for phenylalanine (P323F). To assess the introduction and spread
258 of P323L/F in Nevada, the cumulative frequency of P323, L323 and F323 were plotted
259 from March 6th to June 5th (Fig. 3A). Nevada specimens from the beginning of March
260 (March 6-March 15) show P323 was the predominant variant. As additional specimens
261 were collected and sequenced from March to June there was a shift to a higher frequency
262 of L323 and F323 (Fig. 4A). We performed phylogenetic reconstruction of the Nevada
263 specimens and noted the P323L/F variants on the circular dendrogram with the indicated
264 colors (Fig. 4B). Interestingly, analysis of the Northern Nevada and Southern Nevada
265 specimen showed very different dominant variants (Fig. 4B). In Northern Nevada the
266 F323 was more prevalent, and in Southern Nevada L323 was more prevalent. We used
267 a subsampling of Nextstrain.org data to assess the frequency of P323L/F in the United
268 States and globally during the same time period (March 6th to June 5th). P323 is the
269 predominant variant in Asia, while L323 is more prevalent in other areas of the world and
270 F323 is only appreciably noted in North America (Fig. 4D and 4E).
271
272 **DISCUSSION:**
273 We have developed a novel method which combines specific depletion and
274 enrichment strategies that results in efficient SARS-CoV-2 RNA-seq with high genome
275 coverage and depth. An advantage of this protocol is that it generates sequence data

276 directly from swab specimens without the need to passage the virus in cell culture thereby
277 reducing the handling of infectious material and induction of culture-acquired mutations.
278 Another obstacle in sequencing directly from swab specimens is that most FDA-approved
279 commercially available RNA extraction kits are specifically optimized to recover low
280 amounts of total nucleic acids, include carrier polyA RNA that could be convertible into
281 sequence able molecules, as has been observed previously with RNA-seq of Lassa- or
282 Ebola-positive clinical specimens [27].

283 The workflow incorporates amplification of low-abundance RNA into micrograms
284 of DNA, followed by conversion from a fraction of the DNA into Illumina-compatible
285 sequencing libraries and enrichment of these libraries for SARS-CoV-2 sequences. In
286 addition, during the reverse transcription step a reagent was incorporated to reduce the
287 subsequent amplification of host ribosomal RNA. This approach is robust in that it
288 converts low amounts of RNA into microgram quantities of DNA representative of all the
289 RNA species (aside of rRNA) present in the specimen. This DNA can be stored
290 indefinitely to be interrogated by multiple techniques at a later date. Additionally, RNA
291 amplification is likely less sensitive to low viral abundance compared to RT-PCR. Finally,
292 the use of probes to enrich for coronavirus-specific sequencing library molecules is less
293 sensitive to variants compared to tiling PCR amplicon approaches [28-32].

294 The data herein implicate that early in the pandemic, before the “stay-at-home”
295 order on April 1st, there were multiple introductions of SARS-CoV-2 into the state of
296 Nevada. From April 1st to the beginning of June Nevada experienced a period of semi-
297 isolation, as the casinos and most hotels shut down, tourism and travel to the state
298 essentially stopped. Because of the stay-at-home order and social distancing measures

299 put in place there was less mobility of people within and between states. It is possible that
300 these measures, compounded by potential inherent transmission variability of some viral
301 isolates, influenced the change in the frequency of D614G, clades and P323L/F that we
302 noted during this time period within Nevada. In addition, we also found 379C>A had a
303 high prevalence in our study specimens compared to the subsampling of sequences from
304 the United States and international (supplemental Fig. S4). This is a synonymous
305 mutation in nsp1, hence the biological relevance of this nucleotide variant remains to be
306 elucidated.

307 Of the 14,885 complete SARS-CoV-2 genomes currently available (as of August
308 14, 2020) in NCBI there are only 6 genomes that have the P323F variant (accession
309 number: MT706208, LR860619, MT345877, MT627429, MT810889, MT811171). In this
310 study 62 of the 133 specimens from Northern Nevada contain P323F. That is 46% of
311 specimens from Northern Nevada contain P323F compared to 0.04% of NCBI deposited
312 SARS-CoV-2 isolates. We performed structural modeling of the P323L/F variation and
313 did not find any significant change to the nsp12 confirmation with either P323L or P323F
314 (Supplemental Fig. S5), therefore this variant is most likely a neutral mutation and does
315 not confer either a fitness advantage or disadvantage to transmission or pathogenicity of
316 SARS-CoV-2.

317 We find that the overall trend for D614G in Nevada during this time period was
318 similar to what was observed in other states and internationally, with the exception of
319 within Asia where the D614 allele had originated. We noted that there were differences
320 between Northern Nevada and Southern Nevada. In Northern Nevada clade 20C and
321 F323 were more frequent, while during this same time period in Southern Nevada clade

322 20A and L323 were more prevalent. These data indicate that there were distinct genomic
323 profiles of the SARS-CoV-2 viruses that were circulating in these populations during the
324 initial months of the pandemic while stay-at-home order were in place to help prevent
325 transmission of the virus.
326
327 **CONCLUSIONS:**
328 SARS-COV-2 is currently a global pandemic on a scale that has not been experienced
329 since the 1918 influenza pandemic. Viral genomes accumulate nucleotide variations while
330 passing in the human population and these mutations may confer phenotypic differences
331 including immune response and anti-viral drug resistance. We have developed a protocol
332 to enrich SARS-CoV-2 RNA directly from patient nasal/nasopharyngeal specimens that
333 does not rely on amplicon based sequencing strategies nor the need of passing the virus
334 into tissue culture and reducing the possible introduction of lab-adapted mutations. These
335 specimens were subjected to next generation sequencing (NGS) to determine the
336 phylogenetic relationship of SARS-CoV-2 sequences within Nevada during the early
337 months (March to June) of the pandemic. Our study identified a very rare mutation in the
338 RdRp protein (P323F) found to occur at a very high frequency among patients of Nevada.
339 This may be important for clinical and public health perspectives in understanding the
340 transmission and evolution of SARS-CoV-2 as it circulates in humans.
341
342 **DATA AVAILABILITY:**
343 All sequences are available at bio project:
344 https://www.ncbi.nlm.nih.gov/bioproject/657893. All reported data are deposited and

available at GISAID with the following accession numbers: EPI_ISL_515409, EPI_ISL_515407, EPI_ISL_515408, EPI_ISL_515375, EPI_ISL_515376, EPI_ISL_515373, EPI_ISL_515374, EPI_ISL_515371, EPI_ISL_515372, EPI_ISL_515370, EPI_ISL_515416, EPI_ISL_515417, EPI_ISL_515414, EPI_ISL_515415, EPI_ISL_515379, EPI_ISL_515412, EPI_ISL_515413, EPI_ISL_515377, EPI_ISL_515410, EPI_ISL_515378, EPI_ISL_515411, EPI_ISL_515364, EPI_ISL_515365, EPI_ISL_515362, EPI_ISL_515363, EPI_ISL_515360, EPI_ISL_515361, EPI_ISL_515405, EPI_ISL_515406, EPI_ISL_515403, EPI_ISL_515404, EPI_ISL_515368, EPI_ISL_515401, EPI_ISL_515369, EPI_ISL_515402, EPI_ISL_515366, EPI_ISL_515367, EPI_ISL_515400, EPI_ISL_515353, EPI_ISL_515354, EPI_ISL_515351, EPI_ISL_515352, EPI_ISL_515350, EPI_ISL_515359, EPI_ISL_515357, EPI_ISL_515358, EPI_ISL_515355, EPI_ISL_515356, EPI_ISL_515342, EPI_ISL_515343, EPI_ISL_515340, EPI_ISL_515461, EPI_ISL_515341, EPI_ISL_515462, EPI_ISL_515460, EPI_ISL_515348, EPI_ISL_515349, EPI_ISL_515346, EPI_ISL_515347, EPI_ISL_515344, EPI_ISL_515345, EPI_ISL_515331, EPI_ISL_515298, EPI_ISL_515452, EPI_ISL_515332, EPI_ISL_515299, EPI_ISL_515453, EPI_ISL_515296, EPI_ISL_515450, EPI_ISL_515330, EPI_ISL_515297, EPI_ISL_515451, EPI_ISL_515294, EPI_ISL_515295, EPI_ISL_515293, EPI_ISL_515339, EPI_ISL_515337, EPI_ISL_515458, EPI_ISL_515338, EPI_ISL_515459, EPI_ISL_515335, EPI_ISL_515456, EPI_ISL_515336, EPI_ISL_515457, EPI_ISL_515333, EPI_ISL_515454, EPI_ISL_515334, EPI_ISL_515455, EPI_ISL_515319,

EPI_ISL_515320, EPI_ISL_515441, EPI_ISL_515321, EPI_ISL_515442, EPI_ISL_515440, EPI_ISL_515328, EPI_ISL_515449, EPI_ISL_515329, EPI_ISL_515326, EPI_ISL_515447, EPI_ISL_515327, EPI_ISL_515448, EPI_ISL_515324, EPI_ISL_515445, EPI_ISL_515325, EPI_ISL_515446, EPI_ISL_515322, EPI_ISL_515443, EPI_ISL_515323, EPI_ISL_515444, EPI_ISL_515308, EPI_ISL_515429, EPI_ISL_515309, EPI_ISL_515390, EPI_ISL_515397, EPI_ISL_515430, EPI_ISL_515310, EPI_ISL_515398, EPI_ISL_515431, EPI_ISL_515395, EPI_ISL_515396, EPI_ISL_515393, EPI_ISL_515394, EPI_ISL_515391, EPI_ISL_515392, EPI_ISL_515317, EPI_ISL_515438, EPI_ISL_515318, EPI_ISL_515439, EPI_ISL_515315, EPI_ISL_515436, EPI_ISL_515316, EPI_ISL_515437, EPI_ISL_515313, EPI_ISL_515434, EPI_ISL_515314, EPI_ISL_515435, EPI_ISL_515311, EPI_ISL_515399, EPI_ISL_515432, EPI_ISL_515312, EPI_ISL_515433, EPI_ISL_515418, EPI_ISL_515419, EPI_ISL_515386, EPI_ISL_515387, EPI_ISL_515420, EPI_ISL_515384, EPI_ISL_515385, EPI_ISL_515382, EPI_ISL_515383, EPI_ISL_515380, EPI_ISL_515381, EPI_ISL_515306, EPI_ISL_515427, EPI_ISL_515307, EPI_ISL_515428, EPI_ISL_515304, EPI_ISL_515425, EPI_ISL_515305, EPI_ISL_515426, EPI_ISL_515302, EPI_ISL_515423, EPI_ISL_515303, EPI_ISL_515424, EPI_ISL_515300, EPI_ISL_515388, EPI_ISL_515421, EPI_ISL_515301, EPI_ISL_515389, EPI_ISL_515422

ACKNOWLEDGEMENTS:

We thank the staff at the Nevada State Public Health Lab (NSPHL). This work was supported by NSPHL, University of Nevada, Reno Vice President for Research and Innovation (VPRI), Department of Microbiology & Immunology, UNR School of Medicine, Nevada IDeA Network of Biomedical Research Excellence (INBRE) from the National Institute of General Medical Sciences (GM 103440 and GM 104944) from the National Institutes of Health (NIH). The authors wish to acknowledge the support of Research & Innovation and the Office of Information Technology at the University of Nevada, Reno for computing time on the Pronghorn High-Performance Computing Cluster.

AUTHOR CONTRIBUTION:

PDH: conceptualization, formal analysis, methodology, writing – original draft preparation, writing – review and editing

RLT: conceptualization, formal analysis, methodology, writing – original draft preparation, writing – review and editing

XY: formal analysis

DPA: conceptualization, funding

JRS: formal analysis, review and editing

AG: methodology

EB: specimen procurement and diagnostic testing

HH: specimen procurement and diagnostic testing

MP: specimen procurement and diagnostic testing

MP: conceptualization, specimen procurement, diagnostic testing, formal analysis, project administration, writing – review and editing

CCR: conceptualization, formal analysis, methodology, project administration, funding, writing – original draft preparation, writing – review and editing

SCV: conceptualization, formal analysis, methodology, project administration, funding, writing – review and editing

DECLARATIONS:

Ethics Approval:

Deidentified human specimens (nasal and nasopharyngeal swabs) were used for the extraction of viral RNA all the experiments were done in accordance with guidelines of the University of Nevada, Reno. The University of Nevada, Reno Institutional Review Board (IRB) reviewed this project and determined this study to be EXEMPT FROM IRB REVIEW according to federal regulations and University policy.

The Environmental and Biological Safety committee of the University of Nevada, Reno, approved methods and techniques used in this study.

Competing Interests:

The authors declare that they have no competing interests with the contents of this article.

REFERENCES:

1. Coronaviridae Study Group of the International Committee on Taxonomy of V: **The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2**. *Nat Microbiol* 2020, **5**(4):536-544.
2. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E: **COVID-19, SARS and MERS: are they closely related?** *Clin Microbiol Infect* 2020, **26**(6):729-734.

438 3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang
439 CL *et al*: **A pneumonia outbreak associated with a new coronavirus of**
440 **probable bat origin**. *Nature* 2020, **579**(7798):270-273.
441 4. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, Yang L, Ding C, Zhu X, Lv R *et al*: **Genomic**
442 **characterization and infectivity of a novel SARS-like coronavirus in Chinese**
443 **bats**. *Emerg Microbes Infect* 2018, **7**(1):154.
444 5. de Wit E, van Doremalen N, Falzarano D, Munster VJ: **SARS and MERS: recent**
445 **insights into emerging coronaviruses**. *Nat Rev Microbiol* 2016, **14**(8):523-534.
446 6. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan
447 WW, Cheung MT *et al*: **Coronavirus as a possible cause of severe acute**
448 **respiratory syndrome**. *Lancet* 2003, **361**(9366):1319-1325.
449 7. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA: **Isolation**
450 **of a novel coronavirus from a man with pneumonia in Saudi Arabia**. *N Engl J*
451 *Med* 2012, **367**(19):1814-1820.
452 8. Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, Li PH, Tan SY, Chang Q,
453 Xie JP *et al*: **Epidemiology and cause of severe acute respiratory syndrome**
454 **(SARS) in Guangdong, People's Republic of China, in February, 2003**. *Lancet*
455 2003, **362**(9393):1353-1358.
456 9. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H: **The Architecture of SARS-**
457 **CoV-2 Transcriptome**. *Cell* 2020, **181**(4):914-921 e910.
458 10. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A:
459 **Isolation and Full-Length Genome Characterization of SARS-CoV-2 from**
460 **COVID-19 Cases in Northern Italy**. *J Virol* 2020, **94**(11).

461 11. Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, Halden RU,
462 Hogue BG, Scotch M, Lim ES: **An 81-Nucleotide Deletion in SARS-CoV-2**
463 **ORF7a Identified from Sentinel Surveillance in Arizona (January to March**
464 **2020)**. *J Virol* 2020, **94**(14).
465 12. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei
466 YY *et al*: **A new coronavirus associated with human respiratory disease in**
467 **China**. *Nature* 2020, **579**(7798):265-269.
468 13. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ,
469 Zambon M, Ellis J, Lewis PA, Hiscox JA *et al*: **Characterisation of the**
470 **transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced**
471 **in-frame deletion of the furin-like cleavage site from the spike glycoprotein**.
472 *Genome Med* 2020, **12**(1):68.
473 14. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D: **Structure,**
474 **Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein**. *Cell* 2020,
475 **181**(2):281-292 e286.
476 15. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina**
477 **sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.
478 16. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat*
479 *Methods* 2012, **9**(4):357-359.
480 17. Picard-Toolkit.: **[cited 17 Aug 2020]**. *In: Github [Internet]* 2019, **Available:**
481 **http://broadinstitute.github.io/picard**.
482 18. Garrison E, Marth G: **Haplotype-based variant detection from short-read**
483 **sequencing**. *arXiv [q-bioGN]* 2012, **Available: http://arxiv.org/abs/1207.3907**.

484 19. Garrison E: **Vcflib: A C++ library for parsing and manipulating VCF files**.
485 *Available:* https://githubcom/vcflib/vcflib 2019.
486 20. Li H: **A statistical framework for SNP calling, mutation discovery, association**
487 **mapping and population genetical parameter estimation from sequencing**
488 **data**. *Bioinformatics* 2011, **27**(21):2987-2993.
489 21. Elbe S, Buckland-Merrett G: **Data, disease and diplomacy: GISAID's innovative**
490 **contribution to global health**. *Glob Chall* 2017, **1**(1):33-46.
491 22. Nextstrain.: **Genomic epidemiology of novel coronavirus - Global**
492 **subsampling**. *In: Nextstrainorg [Internet] 15 Jul 2020 [cited 15 Jul 2020]* 2020,
493 **Available: https://nextstrain.org/ncov/global/2020-07-**
494 **15?d=tree&l=clock&legend=closed**.
495 23. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P,
496 Bedford T, Neher RA: **Nextstrain: real-time tracking of pathogen evolution**.
497 *Bioinformatics* 2018, **34**(23):4121-4123.
498 24. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi
499 K, Eng ET, Vatandaslar H, Chait BT *et al*: **Structural Basis for Helicase-**
500 **Polymerase Coupling in the SARS-CoV-2 Replication-Transcription**
501 **Complex**. *Cell* 2020.
502 25. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner
503 N, Giorgi EE, Bhattacharya T, Foley B *et al*: **Tracking Changes in SARS-CoV-2**
504 **Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus**. *Cell*
505 2020, **182**(4):812-827 e819.
506 26. Nextclade: **[cited 17 Aug 2020]**. *Available:* https://cladesnextstrainorg 2020.

507 27. Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R,
508 Stremlau M, Berlin A, Gire SK, England E *et al*: **Enhanced methods for unbiased**
509 **deep sequencing of Lassa and Ebola RNA viruses from clinical and**
510 **biological samples**. *Genome Biol* 2014, **15**(11):519.
511 28. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI: **Virome**
512 **Capture Sequencing Enables Sensitive Viral Diagnosis and Comprehensive**
513 **Virome Analysis**. *mBio* 2015, **6**(5):e01491-01415.
514 29. O'Flaherty BM, Li Y, Tao Y, Paden CR, Queen K, Zhang J, Dinwiddie DL, Gross
515 SM, Schroth GP, Tong S: **Comprehensive viral enrichment enables sensitive**
516 **respiratory virus genomic identification and analysis by next generation**
517 **sequencing**. *Genome Res* 2018, **28**(6):869-877.
518 30. Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, Tong S: **Rapid, Sensitive,**
519 **Full-Genome Sequencing of Severe Acute Respiratory Syndrome**
520 **Coronavirus 2**. *Emerg Infect Dis* 2020, **26**(10).
521 31. Paskey AC, Frey KG, Schroth G, Gross S, Hamilton T, Bishop-Lilly KA:
522 **Enrichment post-library preparation enhances the sensitivity of high-**
523 **throughput sequencing-based detection and characterization of viruses**
524 **from complex samples**. *BMC Genomics* 2019, **20**(1):155.
525 32. Xiao M, Liu X, Ji J, Li M, Li J, Yang L, Sun W, Ren P, Yang G, Zhao J *et al*: **Multiple**
526 **approaches for massively parallel sequencing of SARS-CoV-2 genomes**
527 **directly from clinical samples**. *Genome Med* 2020, **12**(1):57.
528
529

530 **FIGURE LEGENDS:**

531 **Figure 1. Workflow of SARS-CoV-2 genome sequencing and analysis from**
532 **nasopharyngeal patient specimens in Nevada.** (A) RNA was extracted from Nasal or
533 Nasopharygeal (NP) swabs taken from patients in Nevada and first used to determine the
534 presence of SARS-CoV-2 genomes by RT-qPCR. Next generation sequencing (NGS)
535 libraries were prepared from positive specimens, this included steps for ribosomal RNA
536 depletion and SARS-CoV-2 enrichment. Subsequent libraries were pooled and used for
537 whole genome sequencing at the Nevada Genomics Center on the Illumina NextSeq 500
538 instrument. FASTQ files were aligned to the reference genome, and analyzed to
539 determine nucleotide variation and phylogenetic relationship. (B) A total of 200 specimens
540 were sequenced, of which 174 had over 99% coverage of the SARS-CoV-2 genome.
541 This included 133 patient specimens from Northern Nevada, 40 from Southern Nevada
542 and 1 specimen that was re-sequenced. (C) Correlation between RT-qPCR Ct value and
543 the percentage of coverage in the whole genome sequencing after trimming and
544 alignment. (D) Nucleotide variants across the SARS-CoV-2 genome in the 173 specimens
545 from Nevada from March 6th to June 5th.

546

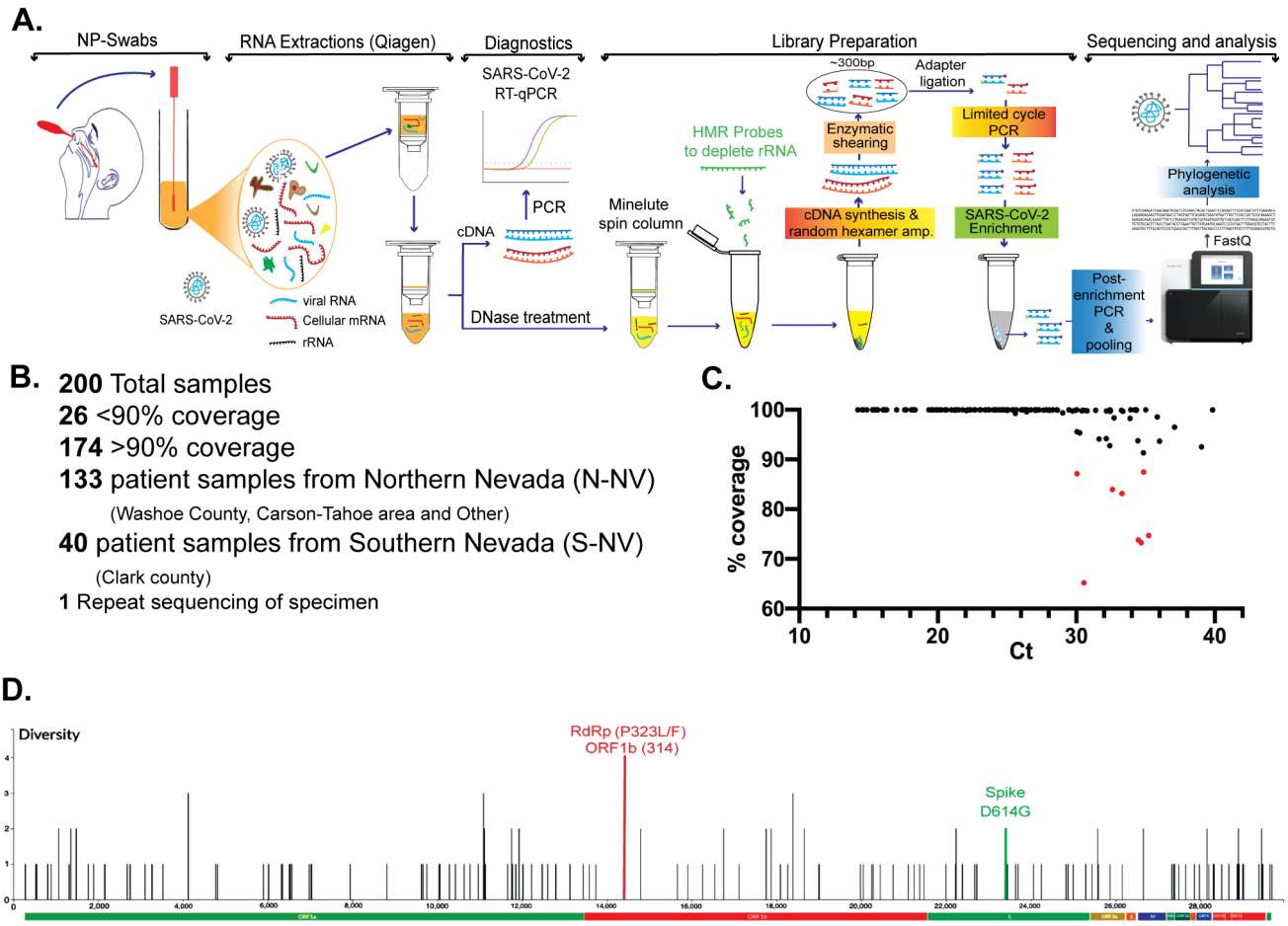
547 **Figure 2. Distribution of D614G in Nevada and comparison with the United States**
548 **and global proportion.** (A) Cumulative frequency of D614G in 173 patient specimens
549 from Nevada from March 6th to June 5th, 2020 (D614 is indicated by teal, G614 is
550 indicated by yellow). Pie charts depict the cumulative proportion up to the indicated time
551 point (March 15, April 1, May 1, June 5). The total number of specimens included at each
552 time point is specified below each pie chart. Effective dates of emergency orders and

553 regulatory responses to SARS-CoV-2 spread in Nevada are indicated on the frequency
554 graph time axis. (B) Proportion of D614G in the United States from March 6th to June 5th,
555 specimens from Nevada are divided in the geographic area that they originated from,
556 Northern Nevada (N-NV) includes 133 specimens from Washoe county, Carson-Tahoe,
557 and other northern counties, and Southern Nevada (S-NV) includes 40 specimens from
558 Clark county. (C) Global proportion of D614G in the shown regions during the same time
559 period from a subsampling of sequences deposited in Nextstrain.org. The size of the pie
560 chart corresponds to the relative specimen number for each region.
561
562 **Figure 3. Distribution of SARS-CoV-2 clades in Nevada.** (A) Cumulative frequency of
563 SARS-CoV-2 clades in 173 patient specimens from Nevada during March 6th to June
564 5th. The five clades are colored 19A (blue), 19B (teal), 20A (green), 20B (yellow) and 20C
565 (orange). Pie charts depict the cumulative proportion up to the indicated time point (March
566 15, April 1, May 1, June 5). The total number of specimens included at each time point is
567 specified below each pie chart. Dates of emergency orders and regulations meant to slow
568 the spread of SARS-CoV-2 in Nevada are indicated on the time scale of the frequency
569 graph. (B) Circular dendrogram depicting clades from Nevada specimens. (C) Pie chart
570 of the clades from northern Nevada (N-NV), southern Nevada (S-NV) and the United
571 States. (D) Pie charts show the proportion of clades from global regions during the same
572 time period from a subsampling of sequences deposited in Nextstrain.org. The size of the
573 pie chart corresponds to the relative specimen number for each region.
574

575 **Figure 4. Distribution of P323L/F (nsp12, RdRp) in Nevada.** (A) Cumulative frequency
576 of P323L/F (nsp12, RdRp) in 173 patient specimens from Nevada during March 6th to
577 June 5th. The amino acid at position 323 is indicated by teal for proline (P), yellow for
578 leucine (L) and blue for phenylalanine (F). Pie charts depict the cumulative proportion up
579 to the indicated time point (March 15, April 1, May 1, June 5). The total number of
580 specimens included at each time point is specified below each pie chart. Dates of
581 emergency orders and regulations meant to slow the spread of SARS-CoV-2 in Nevada
582 are indicated on the time scale of the frequency graph. (B) Circular dendrogram
583 representing the distribution of amino acid change at residue 323 of nsp12 from a global
584 subsampling of sequences deposited in Nextstrain.org from March 6th to June 5th, the
585 larger dots indicate specimens from Nevada. (C) Pie chart indicating the ratio of P/L/F in
586 Northern NV and Southern NV specimens from this study. (D) Proportion of P323L/F from
587 a subsampling of sequences deposited in Nextstrain.org for the United States and (E)
588 global regions from March 6th to June 5th. The size of the pie chart corresponds to the
589 relative specimen number for each region.
590
591
592
593
594
595
596
597

598 **Figure 1:**

599



600

601

602

603

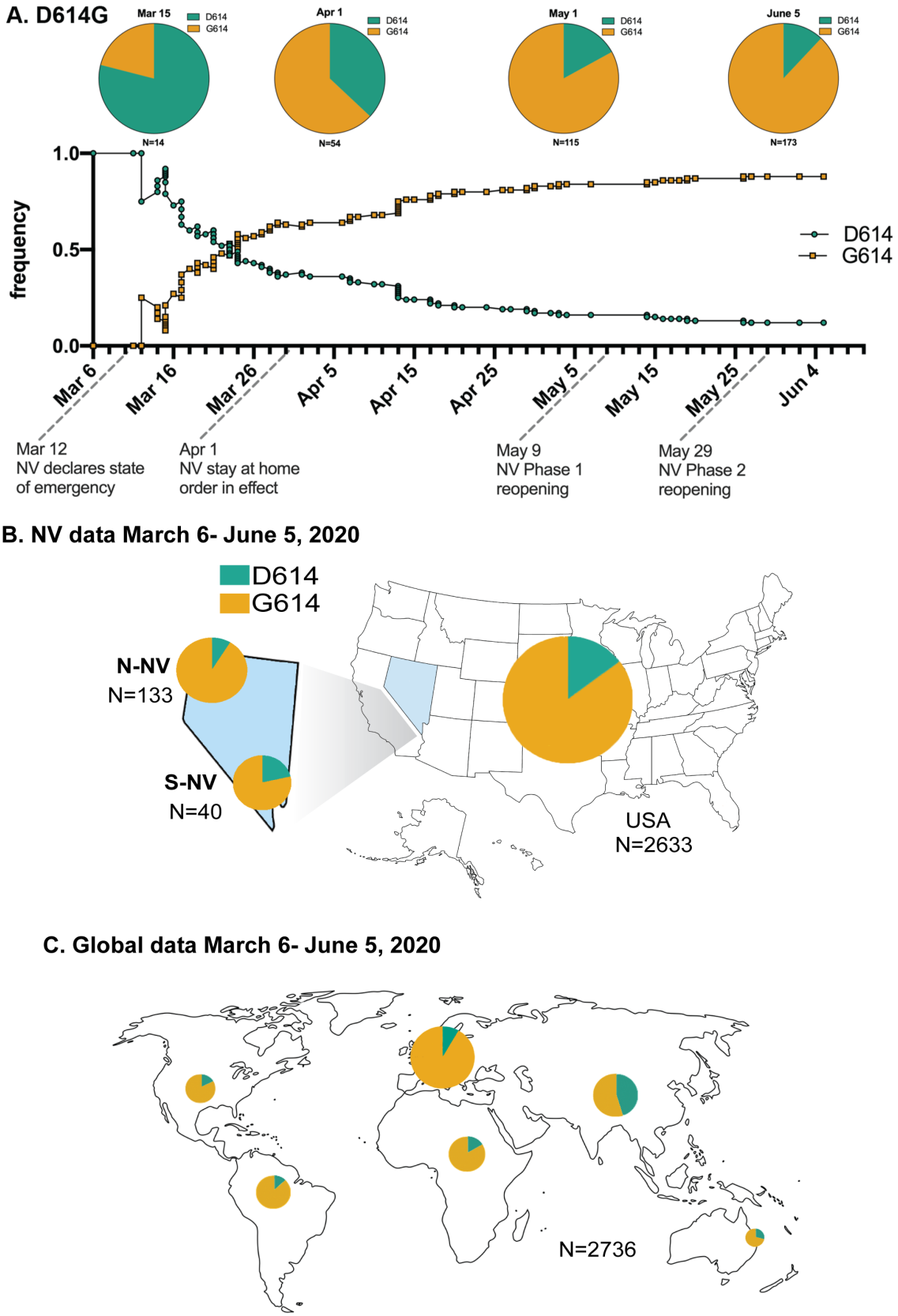
604

605

606

607

Figure 2:


A. D614G
Mar 15
Apr 1
May 1
June 5
D614
G614
N=14
N=54
N=115
N=173
frequency
1.0
0.5
0.0
Mar 6
Mar 16
Mar 26
Apr 5
Apr 15
Apr 25
May 5
May 15
May 25
Jun 4
Mar 12
NV declares state of emergency
Apr 1
NV stay at home order in effect
May 9
NV Phase 1 reopening
May 29
NV Phase 2 reopening
B. NV data March 6- June 5, 2020
N-NV
N=133
S-NV
N=40
USA
N=2633
C. Global data March 6- June 5, 2020
N=2736

631 **Figure 3:**

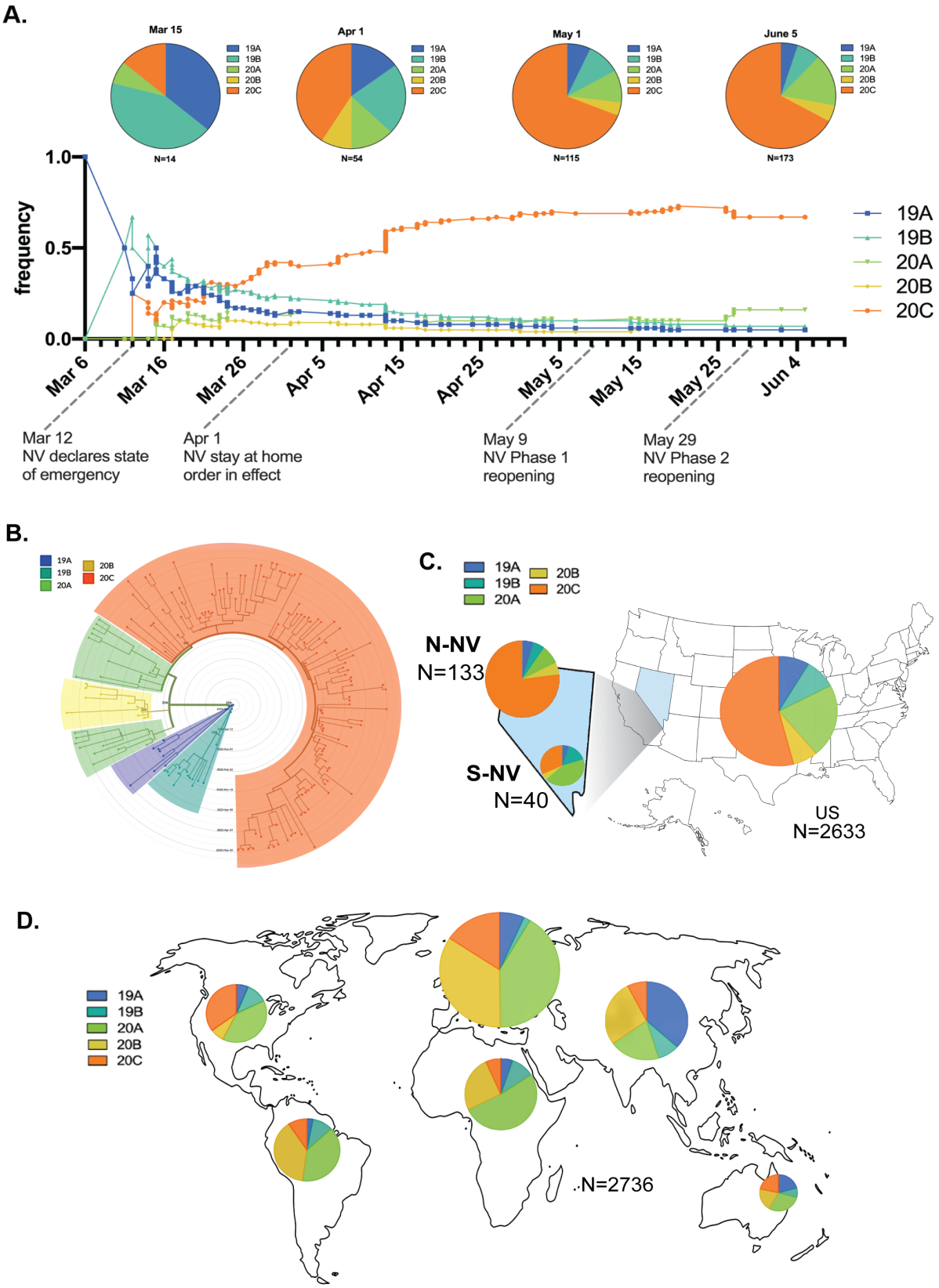
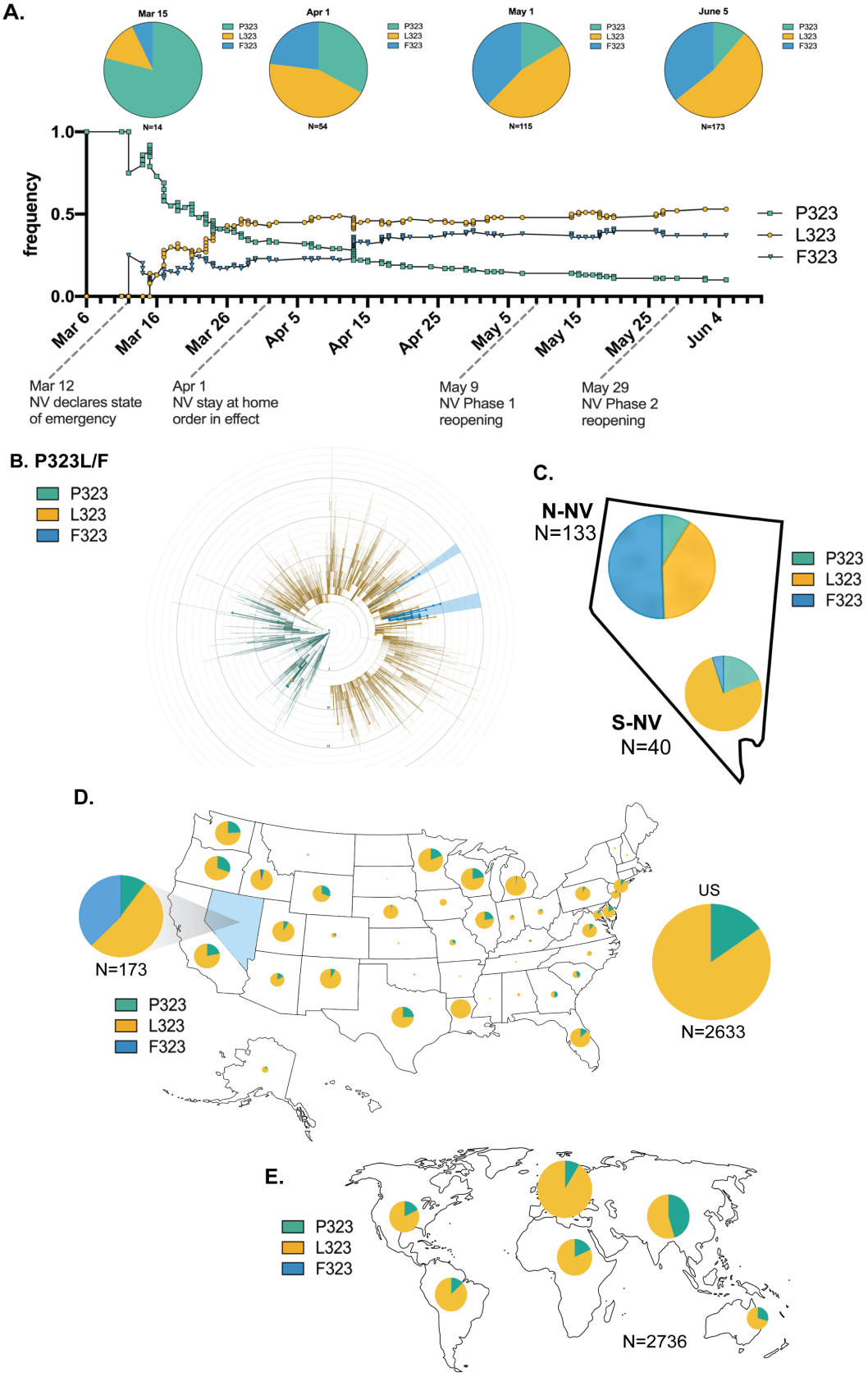

A.
Mar 15
Apr 1
May 1
June 5
19A
19B
20A
20B
20C
N=14
N=54
N=115
N=173
1.0
0.5
0.0
frequency
Mar 6
Mar 16
Mar 26
Apr 5
Apr 15
Apr 25
May 5
May 15
May 25
Jun 4
Mar 12
NV declares state of emergency
Apr 1
NV stay at home order in effect
May 9
NV Phase 1 reopening
May 29
NV Phase 2 reopening
B.
C.
N-NV
N=133
S-NV
N=40
US
N=2633
D.
..N=2736

Figure 4:


A.
Mar 15
Apr 1
May 1
June 5
P323
L323
F323
N=14
N=54
N=115
N=173
frequency
1.0
0.5
0.0
Mar 6
Mar 16
Mar 26
Apr 5
Apr 15
Apr 25
May 5
May 15
May 25
Jun 4
Mar 12
NV declares state of emergency
Apr 1
NV stay at home order in effect
May 9
NV Phase 1 reopening
May 29
NV Phase 2 reopening
B. P323L/F
C.
N-NV
N=133
S-NV
N=40
D.
N=173
US
N=2633
E.
N=2736

SUPPLEMENTARY INFORMATION

Figure S1: Alignment of SARS-CoV-2 sequences from NV patient's specimens.

NCBI Multiple Sequence Alignment Viewer, Version 1.15.0

Sequence ID		Start	End	Organism
Consensus	(+)	1	29,906	
NC_045512.2	(+)	1	29,903	SARS-CoV-2
A0196	(+)	1	29,897	
A0140	(+)	1	29,878	
A0082	(+)	1	29,903	
A0115	(+)	1	29,899	
A0120	(+)	1	29,900	
A0010	(+)	1	29,894	
A0080	(+)	1	29,902	
A0008	(+)	1	29,900	
A0113	(+)	1	29,866	
A0154	(+)	1	29,899	
A0044	(+)	1	29,900	
A0028	(+)	1	29,900	
A0192	(+)	1	29,899	
A0045	(+)	1	29,899	
A0034	(+)	1	29,894	
A0195	(+)	1	29,897	
A0021	(+)	1	29,897	
A0016	(+)	1	29,901	
A0193	(+)	1	29,900	
A0059	(+)	1	29,903	
A0088	(+)	1	29,900	
A0070	(+)	1	29,896	
A0103	(+)	1	29,898	
A0156	(+)	1	29,900	
A0149	(+)	1	29,899	
A0052	(+)	1	29,901	
A0048	(+)	1	29,901	
A0171	(+)	1	29,897	
A0208	(+)	1	29,872	
A0099	(+)	1	29,897	
A0201	(+)	1	29,900	
A0068	(+)	1	29,899	
A0065	(+)	1	29,886	
A0093	(+)	1	29,900	
A0187	(+)	1	29,900	
A0033	(+)	1	29,899	
A0089	(+)	1	29,903	
A0064	(+)	1	29,899	
A0203	(+)	1	29,897	
A0047	(+)	1	29,897	
A0085	(+)	1	29,900	
A0141	(+)	1	29,900	
A0189	(+)	1	29,899	
A0199	(+)	1	29,900	
A0194	(+)	1	29,901	
A0097	(+)	1	29,897	
A0017	(+)	1	29,866	
A0018	(+)	1	29,894	
A0057	(+)	1	29,900	
A0014	(+)	1	29,901	
A0078	(+)	1	29,899	
A0181	(+)	1	29,900	
A0084	(+)	1	29,901	
A0122	(+)	1	29,897	
A0118	(+)	1	29,899	
A0138	(+)	1	29,875	
A0175	(+)	1	29,900	
A0069	(+)	1	29,900	
A0090	(+)	1	29,862	
A0015	(+)	1	29,897	
A0013	(+)	1	29,897	
A0145	(+)	1	29,897	
A0020	(+)	1	29,900	
A0137	(+)	1	29,899	
A0188	(+)	1	29,903	
A0030	(+)	1	29,899	
A0123	(+)	1	29,899	
A0205	(+)	1	29,899	
A0091	(+)	1	29,903	
A0022	(+)	1	29,900	
A0180	(+)	1	29,897	
A0111	(+)	1	29,900	
A0051	(+)	1	29,897	
A0039	(+)	1	29,872	
A0098	(+)	1	29,899	
A0121	(+)	1	29,897	
A0200	(+)	1	29,900	
A0076	(+)	1	29,899	
A0007	(+)	1	29,901	
A0025	(+)	1	29,900	
A0032	(+)	1	29,899	
A0026	(+)	1	29,900	
A0019	(+)	1	29,900	
A0095	(+)	1	29,899	
A0042	(+)	1	29,902	
A0005	(+)	1	29,897	
A0147	(+)	1	29,900	
A0092	(+)	1	29,899	
A0119	(+)	1	29,897	
A0136	(+)	1	29,897	
A0056	(+)	1	29,897	
A0031	(+)	1	29,899	
A0182	(+)	1	29,897	
A0046	(+)	1	29,899	
A0071	(+)	1	29,897	
A0117	(+)	1	29,899	
A0176	(+)	1	29,899	
A0146	(+)	1	29,900	
A0183	(+)	1	29,897	
A0086	(+)	1	29,900	
A0079	(+)	1	29,901	
A0131	(+)	1	29,900	
A0161	(+)	1	29,900	
A0153	(+)	1	29,899	
A0049	(+)	1	29,899	
A0036	(+)	1	29,901	
A0172	(+)	1	29,900	
A0165	(+)	1	29,900	
A0162	(+)	1	29,900	
A0116	(+)	1	29,900	
A0190	(+)	1	29,903	
A0087	(+)	1	29,901	
A0155	(+)	1	29,900	
A0198	(+)	1	29,899	
A0041	(+)	1	29,899	
A0151	(+)	1	29,901	
A0058	(+)	1	29,884	
A0174	(+)	1	29,899	
A0173	(+)	1	29,900	
A0206	(+)	1	29,899	
A0129	(+)	1	29,900	
A0105	(+)	1	29,897	
A0054	(+)	1	29,899	
A0075	(+)	1	29,871	
A0073	(+)	1	29,897	
A0185	(+)	1	29,897	
A0114	(+)	1	29,897	
A0164	(+)	1	29,884	
A0110	(+)	1	29,896	
A0202	(+)	1	29,897	
A0024	(+)	1	29,897	
A0004	(+)	1	29,899	
A0187	(+)	1	29,891	
A0197	(+)	1	29,897	
A0152	(+)	1	29,883	
A0178	(+)	1	29,852	
A0106	(+)	1	29,889	
A0006	(+)	1	29,899	
A0167	(+)	1	29,889	
A0210	(+)	1	29,872	
A0060	(+)	1	29,899	
A0157	(+)	1	29,849	
A0163	(+)	1	29,893	
A0148	(+)	1	29,868	
A0133	(+)	1	29,906	
A0144	(+)	1	29,897	
A0160	(+)	1	29,855	
A0055	(+)	1	29,897	
A0027	(+)	1	29,862	
A0062	(+)	1	29,897	
A0094	(+)	1	29,897	
A0139	(+)	1	29,897	
A0081	(+)	1	29,854	
A0104	(+)	1	29,642	
A0108	(+)	1	29,897	
A0066	(+)	1	29,667	
A0083	(+)	1	29,843	
A0023	(+)	1	29,875	
A0143	(+)	1	29,897	
A0009	(+)	1	29,864	
A0204	(+)	1	29,902	
A0132	(+)	1	29,814	
A0040	(+)	1	29,744	
A0063	(+)	1	29,817	
A0053	(+)	1	29,745	
A0043	(+)	1	29,740	
A0124	(+)	1	29,792	
A0112	(+)	1	29,825	
A0134	(+)	1	29,688	
A0107	(+)	1	29,663	
A0207	(+)	1	29,806	
A0061	(+)	1	29,782	
A0177	(+)	1	29,771	
A0003	(+)	1	29,688	Samples with >90% coverage Not Included in the Analysis
A0101	(+)	1	29,884	
A0072	(+)	1	29,631	
A0102	(+)	1	29,761	
A0029	(+)	1	29,757	
A0067	(+)	1	29,757	
A0109	(+)	1	29,663	
A0209	(+)	1	29,694	
A0179	(+)	1	29,898	
A0035	(+)	1	29,748	
A0184	(+)	1	29,878	
A0142	(+)	1	29,472	
A0191	(+)	1	29,821	

Figure S2. Dendrogram of Nevada specimens in context of other sequenced specimens. (a) Nucleotide mutation clock from Nextstrain.org with the SARS-CoV-2 genome isolated from Washington (USA/WA 1/2020) on January 24th and first specimen from Nevada (A0004) on March 5th indicated in red. (b) Circular dendrogram of Nevada specimens from March 6th to June 5th positioned within a subsample of global sequences from Nextstrain.org during the same time period, the larger dots indicate specimens from Nevada. The five clades are colored 19A (blue), 19B (teal), 20A (green), 20B (yellow) and 20C (orange).

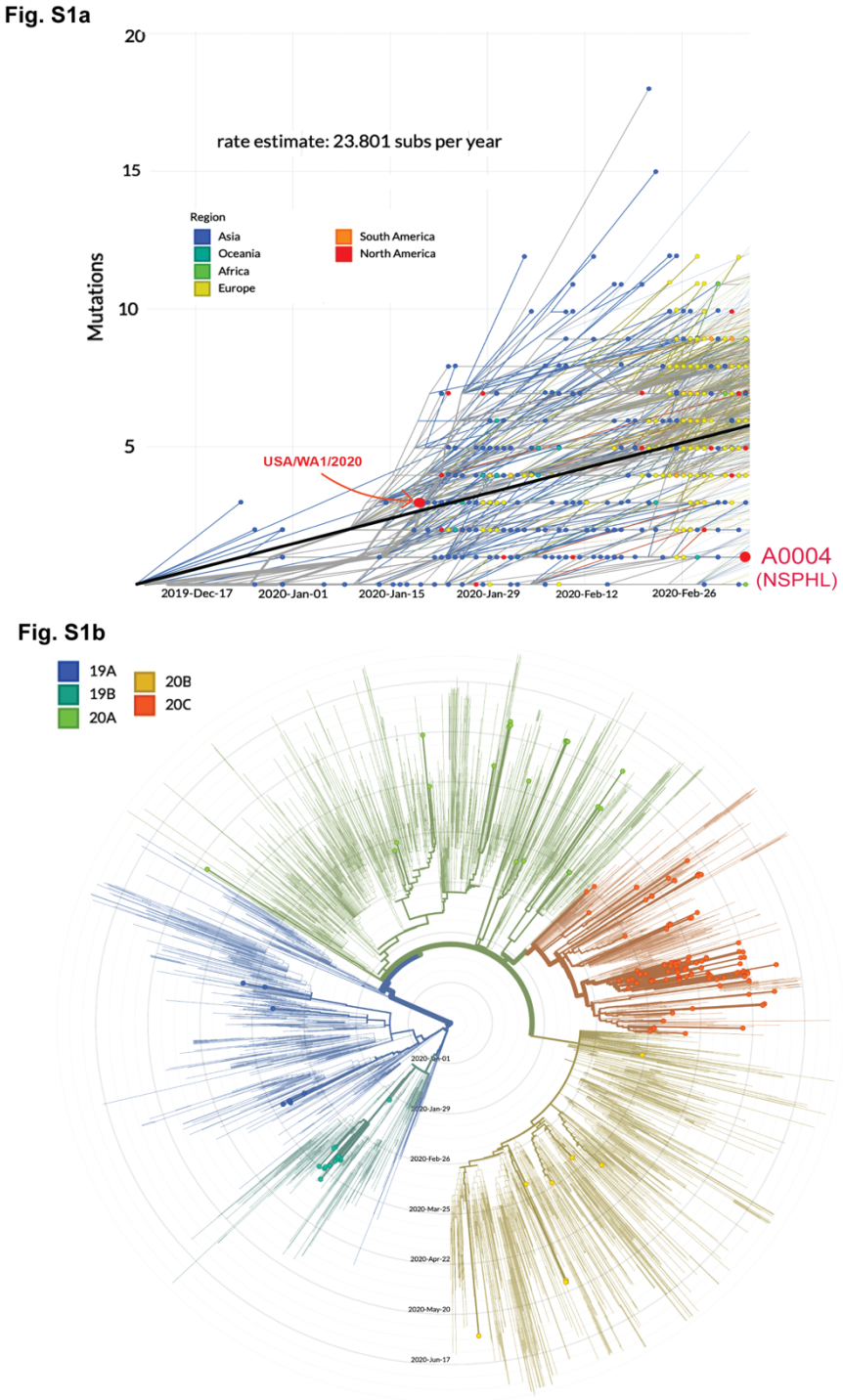


Figure S3. Global distribution of clades from March 6th to June 5th. Pie charts depict the proportion of the clades within (a) four main areas in Nevada, these include 91 specimens from Washoe county (upper left), 23 specimens from Carson-Tahoe (middle left), 40 specimens from Clark county (bottom right), and 19 from rural Nevada (middle). The five clades are colored 19A (blue), 19B (teal), 20A (green), 20B (yellow) and 20C (orange). (b) Pie chart of the clades in each indicated country from March 6th to June 5th generated from a subsampling of sequences deposited in Nextstrain.org. The size of the pie chart corresponds to the relative specimen number for each region.

Fig. S2a

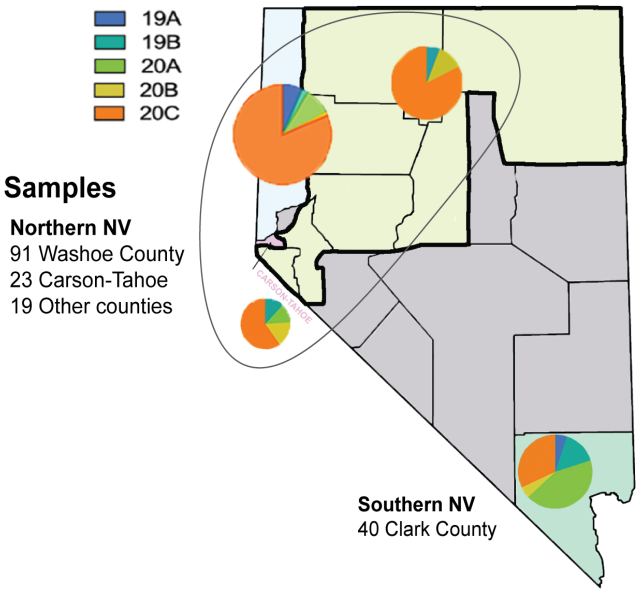


Fig. S2b

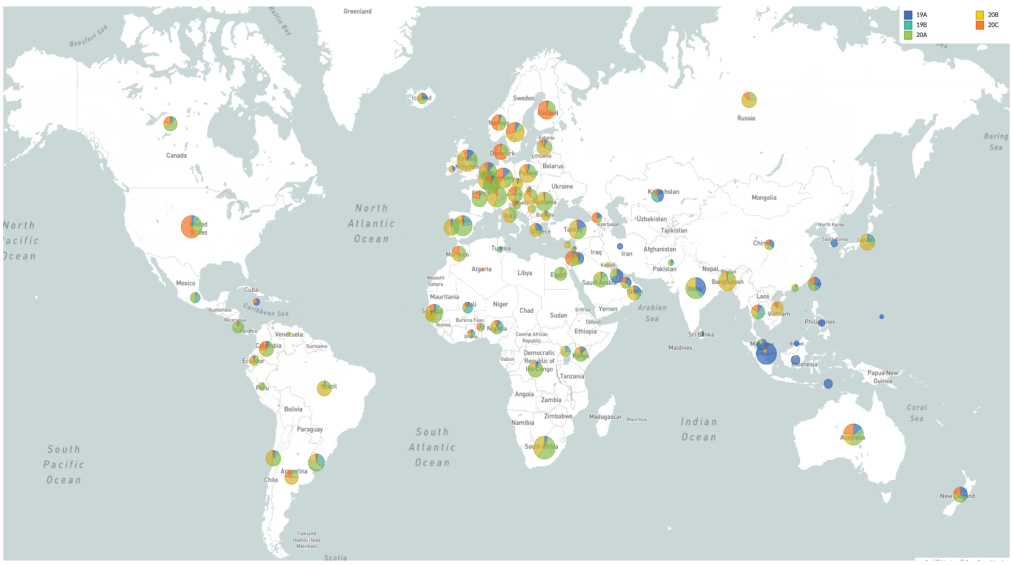


Figure S4. Distribution of nucleotide variant 379C>A. (a) Green line at the far-left end of the genome denoted nucleotide position 379 of nsp1. (b) Circular dendrogram of global subsample of sequences from Nextstrain.org with NV specimens indicated by larger dots. (c) Pie chart indicating the proportion of sequences with either the cytidine (C) or adenosine (A) at position 379 from the Nevada specimens. Subsample of sequences from Nextstrain.org were used to generate the proportion of 379C>A in (d) the indicated states within the U.S. and (e) internationally. The size of the pie chart corresponds to the relative specimen number for each region.

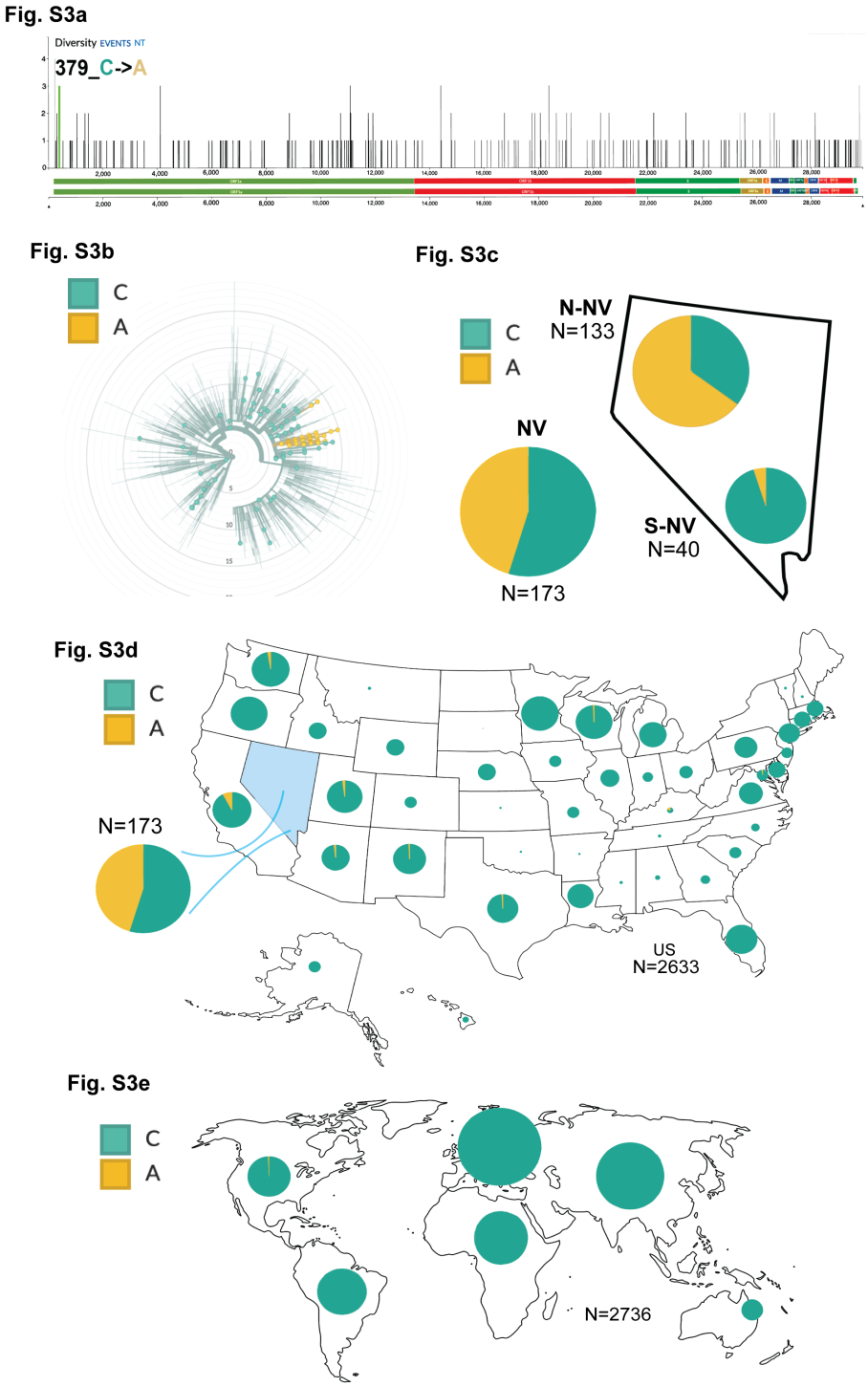
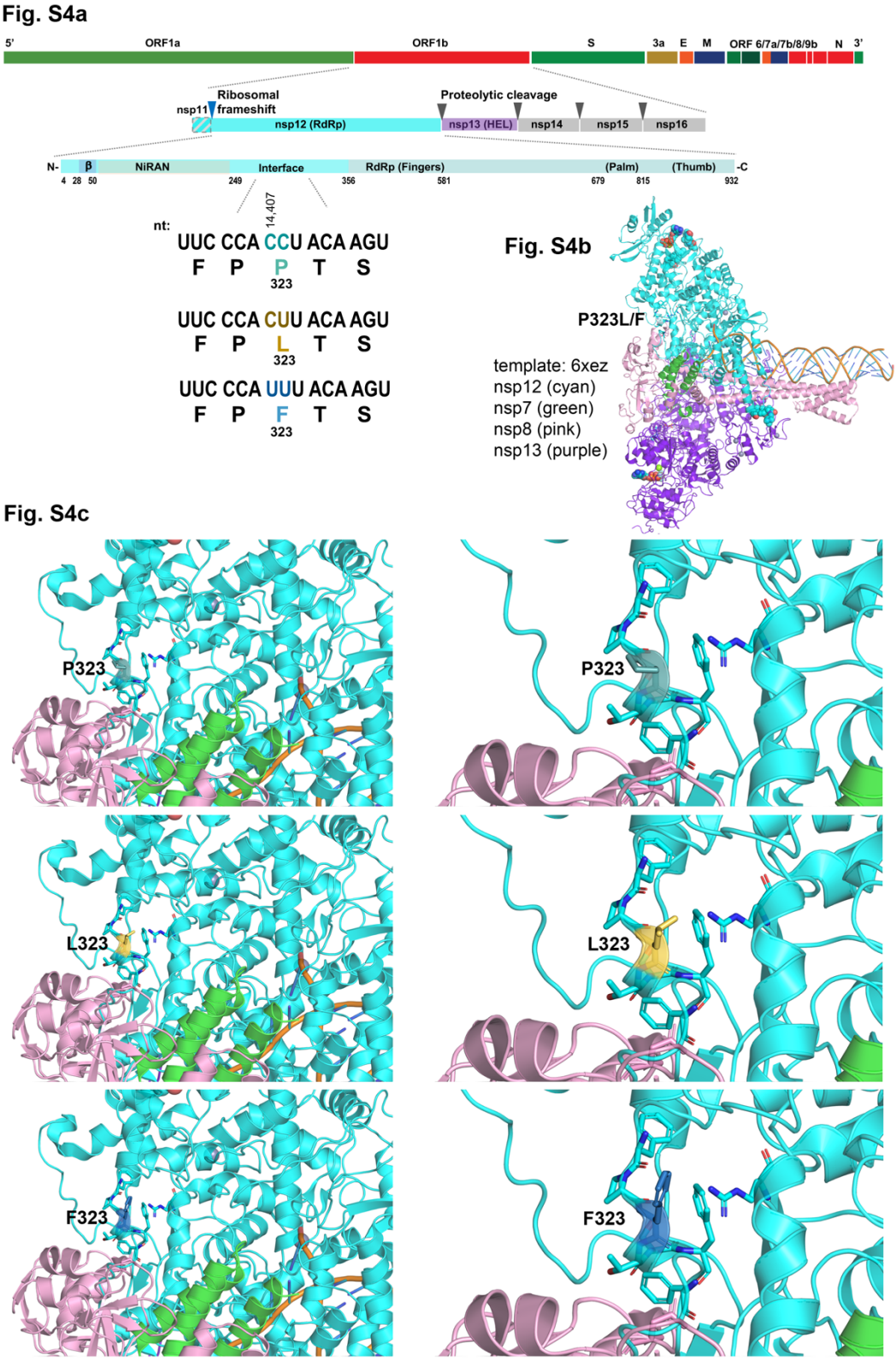


Figure S5. Structure of SARS-CoV-2 nsp12 (RdRp) P323L/F. (a) Diagram depicting ORF1b genomic location and encoded proteins. Below the linear protein schematic of nsp12 specific nucleotide variants at position 14,407 and 14,408 and the resulting amino acid changes are indicated. (b) SARS-CoV-2 replicase complex



121 modeled from 6XEZ template. This model includes nsp12, nsp7, nsp8, nsp13, ligands
122 (ZN, Mg), the template and product strand of RNA. P323L/F within the interface domain
123 of nsp12 is located at the left side of the complex. (c) Model nsp12 showing P323L/F
124 within the interface domain. Residue 323 is shown with either P, L or F and amino
125 acids with side chains within 5 Å of residue 323 are depicted as sticks in the cartoon
126 model.
127