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*], David P. AuCoin[3,4], Joel R. Sevinsky[6], Yanji Xu[4,5], 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:

Patients with signs of COVID-19 were tested with CDC approved diagnostic RT-PCR for SARS-CoV-2 using RNA extracted from nasopharyngeal/nasal swabs. In order to determine the variants of SARS-CoV-2 circulating in the state of Nevada, 200 patient specimens from positively identified cases were sequenced through our robust protocol for sequencing SARS-CoV-2 genomes from the nasopharyngeal or nasal swabs. This protocol enabled the identification of specific nucleotide variants including those coding for D614G and clades defining mutations. Additionally, these sequences were used for determining the phylogenetic relationships of SARS-CoV-2 genomes of public health importance occurring in the state of Nevada. Our study reports the occurrence 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-Hu-1). This 323F variant is found at a very high frequency (46% of the tested specimen) in Northern Nevada, possibly because the virus accumulated this mutation while circulating in the community and the shelter in place orders restricted the introduction and spread of other variants into this region. Structural modeling of the RdRp with P323F variant did not show any significant difference in protein conformation, but the phenotypic effect is unknown and an area of active investigation. In conclusion, 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.

50 **IMPORTANCE:**

51 SARS-COV-2 genomes accumulate nucleotide mutations while passing in the
52 human population and these mutations may confer phenotypic differences including
53 altered immune response and anti-viral drug resistance. We developed a robust workflow
54 to sequence SARS-CoV-2 directly from the nasal/nasopharyngeal swabs containing even
55 a very low viral loads (>35 Ct value samples). Our protocol does not rely on amplicon
56 based sequencing strategies nor the need of passing the virus into tissue culture thus
57 reduces the possibility of an introduction of laboratory-adapted mutations. Sequences of
58 SARS-CoV-2 from the patients of the state of Nevada during early months of the
59 pandemic identified a rare mutation in the RdRp protein (P323F). This mutation occurred
60 at a very high frequency in the variants of SARS-CoV-2 circulating Northern Nevada.
61 Identification of such variants is important for clinical and public health perspectives in
62 understanding transmission mediated evolution of SARS-CoV-2 variants and their
63 implications on therapeutics and diagnostics.

64

65

66

67

68

69

70

71

72

73 **INTRODUCTION:**

74 Severe Acute Respiratory Syndrome coronavirus 2 (SARS-CoV-2), the cause of
75 coronavirus disease 2019 (COVID-19), was first identified and reported in December
76 2019 in Wuhan, Hubei province, China (1-3). RNA sequencing and phylogenetic analysis
77 of specimens taken during the initial outbreak in Wuhan determined that the virus is most
78 closely related (89.1% nucleotide similarity) to a group of SARS-like coronaviruses (genus
79 Betacoribavirus, subgenus Sarbecovirus) which had previously been identify in bats in
80 China (2, 4). Coronaviruses have a recent history as emerging infections, first SARS-CoV
81 in 2002-2003, and Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012,
82 both zoonotic infections that cause severe respiratory illness in humans (2, 5-8) [2,5-8].
83 Unlike SARS-CoV and MERS-CoV which displayed limited global spread, SARS-CoV-2
84 has spread around the world within a few months. There are specific characteristics of
85 SARS-CoV-2 which have facilitated the transmission, including infections that result in
86 asymptomatic or mild disease, allowing for under-characterized transmission.

87 SARS-CoV-2 is an enveloped, positive single-stranded RNA virus. Detection of
88 SARS-CoV-2 in patients has primarily occurred using RT-qPCR to detect viral RNA from
89 respiratory specimens (primarily nasal and nasopharyngeal swabs). While RT-PCR
90 results can be quantified through determination of a cycle threshold (Ct) value for each
91 sample, it does not yield sequence data leading to the description of genomic variants.
92 To further study of such variants, and to better understand the epidemiology of the virus
93 in the state of Nevada, we developed a workflow that allowed us to sequence SARS-CoV-
94 2 genomic RNA from patient swabs containing a broad range of viral loads. Of the
95 sequences of SARS-CoV-2 currently submitted to common database (GenBank and

96 GISAID), several were obtained after the virus had been passed in Vero cells (9, 10) and
97 others came directly from patient specimens (11, 12). Certain data have suggested a
98 potential for lab acquired mutations following passage in cell culture (9, 13). Specifically,
99 a report of SARS-CoV-2 passage in Vero cells which resulted in a spontaneous 9 amino
100 acid deletion within the spike (S) protein that overlaps with the furin cleavage site (13).
101 The loss of this site is suggested to increase the viral entry into Vero cells (14). For both
102 research and epidemiological purposes, sequencing of SARS-CoV-2 directly from patient
103 specimens not only reduces the possibility of laboratory acquired mutations following
104 passage in cell culture but also reduces the time that would be spent growing the virus
105 from the patient specimens and subsequently also reduces handling larger amounts of
106 infectious virus. Additionally, one of the goals in developing an optimized SARS-CoV-2
107 NGS protocol was to be able to generate adequate depth of coverage of the viral genome
108 while minimizing the sequencing of non-viral RNA which would allow for more specimens
109 to be multiplexed together during sequencing.

110 Our workflow employs a combination of RNA amplification, conversion into
111 Illumina-compatible sequencing libraries and enrichment of SARS-CoV-2 library
112 molecules prior to sequencing. Using this novel methodology, we sequenced SARS-
113 CoV-2 from a total of 200 patient specimens collected over a three-month period
114 originating from Nevada. Of the 200 selected, 173 were sequenced with enough quality
115 to be used for determining SARS-CoV-2 nucleotide variants to perform further
116 phylogenetic analysis and study the viral epidemiology within the state of Nevada.
117 Analysis of the data suggests a specific epidemiological course for the local epidemic
118 within Northern Nevada. This was characterized by an initial observation of variants

closely resembling isolates originating directly from China or Europe. Subsequent to government-mandated period of restrictions on business and social activity, we observed that a viral isolate not seen elsewhere in the world emerged within Northern Nevada cases (nucleotide 14,407 and 14,408). This isolate contains an amino acid change in residue P323L/F of RdRp (nsp12). Furthermore, we found that sampled viral isolates in Southern Nevada, unlike those in Northern Nevada, closely resembled the makeup of the United States in general.

RESULTS:

RNA-seq workflow and assembly of SARS-CoV-2 genomes.

A total of 200 SARS-CoV-2 positive specimens collected in Nevada from March 6 to June 5 were randomly selected to have their viral genomes sequenced for variant analysis and subsequent epidemiological studies (Fig 1A and Materials & Methods). Of the sequenced specimens, 173 had >90% coverage and sufficient depth to accurately 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 and surrounding cities). 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

142 figure 1D, along with the genomic areas that were assessed for change in frequency
143 corresponding to amino acids D614G, P323L/F and nucleotide 379.

144 During the sequencing analysis we also examined the correlation between Ct
145 values from the diagnostic RT-PCR and percentage coverage of the viral genome to
146 determine the performance and robustness of our sequencing method in relation to
147 available viral RNA in a specimen of a given Ct (Fig. 1C). In general, a Ct value less than
148 31 resulted in at least 90% coverage to the SARS-CoV-2 genome. Importantly, our in-
149 house developed method for viral genome enrichment and sequencing directly from the
150 patient's specimens (nasal and nasopharyngeal swabs) was robust and yielded
151 sequences covering over 90% of the genome even in samples having very high Ct (~40)
152 of viral genome detection. This is highly significant and shows the power of our workflow
153 in sequencing of SARS-CoV-2 genome from a spectrum of samples including the ones
154 having inadequate amounts of specimen (due to the variability in collection) or lower viral
155 loads in nasal secretions. Consequently, our sequencing protocol avoids any molecular
156 epidemiological bias, which may get acquired through cell culture-based amplification
157 especially in those specimens with high Ct (low viral load) as our method eliminates the
158 need of virus culture.

159 **Prevalence of amino acid variant D614G of SARS-CoV-2 spike protein in specimens**
160 **collected in Nevada.**

161 Earlier studies have revealed the emergence, spread and potential importance of
162 an alteration, D614G (genomic change at 23403A>G), of the spike protein (15). This
163 missense mutation has become a clade-distinguishing locus that differentiates viral
164 isolates originating in Asia from those that have emerged from Europe. A total of 173

165 cases were analyzed to determine the number and relative proportion of the specimens
166 which carried the D614G spike protein variant in Nevada. The cumulative frequency for
167 D614 and G614 were plotted from March 6 to June 5 (Fig. 2A). Specimens from the
168 beginning of March represent the earliest known cases in Nevada, and of the 14
169 specimens sequenced during this time period (March 6-March 15) D614 was the
170 predominant variant. This shifted from March to June with an increasing frequency of the
171 G614 allele. The trend for specimens originating from either Northern Nevada (N-NV) and
172 Southern Nevada (S-NV) both show a higher frequency of G614 (Fig. 2B). We used a
173 subsampling of sequence data from Nextstrain.org to assess the frequency of D614G in
174 the United States and globally during the same time period (March 6 to June 5) (Fig. 2B
175 and 2C). The global trend by continent of D614G is also similar, with G614 at a higher
176 frequency, the one noted exception is in Asia, where D614 and G614 continue to exist in
177 equal proportions (Fig. 2C).

178 **Frequency of SARS-CoV-2 clades in Nevada.**

179 Worldwide, there are currently 5 main clades (19A, 19B, 20A, 20B, 20C) of SARS-
180 CoV-2 differentiated based on specific nucleotide profiles in the Year-letter scheme of
181 https://clades.nextstrain.org. Clade 19A and 19B are defined by C8782T and T28144C,
182 respectively. 20A is a derivative of 19A and contains mutations C3037T, C14408T and
183 A23403G (resulting in D614G). 20B is defined by mutations G28881A, G28882A and
184 G28883C, and 20C contains C1059T and G25563T (16).

185 To assess the introduction and spread of the clades in Nevada the cumulative
186 frequency for the clades were plotted from March 6 to June 5 (Fig. 3A). The earliest
187 sequenced specimens from Nevada were collected in the beginning of March (March 6-

188 March 15) and are predominantly from clades 19A and 19B. Additional sequenced
189 specimens collected from March to June revealed a shift to a higher frequency of 20C
190 (Fig. 3A). We performed phylogenetic reconstruction of the Nevada specimens and
191 differentiated the clades on the circular dendrogram by color (Fig. 3B). There were
192 discordant trends in the dominant clade for specimens originating from either Northern
193 Nevada (N-NV) and Southern Nevada (S-NV) (Fig. 3C and 3D). Specimens from
194 Northern Nevada (Washoe County, Carson-Tahoe, and other counties) showed a
195 prevalence of 20C, while the Southern Nevada specimens from Clark County had a larger
196 proportion of 20A (Fig. 3C and Supplemental Fig. S3a). We used a subsampling of
197 Nextstrain.org data to assess the frequency of clades in the United States and globally
198 by continent during the same time period (March 6 to June 5). The dominant clade in the
199 United States was 20C, similar to the frequency seen in the total Nevada samples (N-NV
200 and S-NV) (Fig. 3A and 3C). The global clade distributions were variable in areas outside
201 of Asia while clades 19A and 19B are noted to be more prevalent in Asia. (Fig. 3D).

202 **Prevalence of amino acid variant P323L/F of SARS-CoV-2 nsp12 (RdRp) in Nevada.**

203 Analysis of sequencing data revealed a novel observation for our specimens at
204 bases 14,407 and 14,408 which results in a change at residue 323 in nsp12 (RdRp). For
205 the Wuhan isolate at 14,407 and 14,408 there is CC for proline (P), the variants have CT
206 for leucine (P323L) and TT for phenylalanine (P323F). To assess the introduction and
207 spread of P323L/F in Nevada, the cumulative frequency of P323, L323 and F323 were
208 plotted from March 6 to June 5 (Fig. 3A). Nevada specimens from the beginning of March
209 (March 6-March 15) show P323 was the predominant variant. As additional specimens
210 were collected and sequenced from March to June there was a shift to a higher frequency

211 of L323 and F323 (Fig. 4A). We performed phylogenetic reconstruction of the Nevada
212 specimens and noted the P323L/F variants on the circular dendrogram with the indicated
213 colors (Fig. 4B). Interestingly, analysis of the Northern Nevada and Southern Nevada
214 specimen showed very different dominant variants (Fig. 4B). In Northern Nevada the
215 F323 was more prevalent, while in Southern Nevada L323 was more prevalent. We used
216 a subsampling of Nextstrain.org data to assess the frequency of P323L/F in the United
217 States and globally during the same time period (March 6 to June 5). P323 is the
218 predominant variant in Asia, while L323 is more prevalent in other areas of the world and
219 F323 is only appreciably noted in North America (Fig. 4D and 4E).

220 While we are investigating the phenotypic significance of this predominant variant
221 of RdRp, we performed in-silico structural modeling of RdRp to determine the spatio-
222 temporal location of this 323aa on RdRp in complex with its accessory proteins, nsp7,
223 nsp8 and nsp3. Our data showed the location of 323aa in the interface domain of RdRp
224 and variation of P323 to L or F did not significantly change the conformation of the protein
225 (Fig. 5). This suggested that P323 variants are most likely causing a neutral mutation in
226 the RdRp without conferring any fitness advantage or disadvantage in transmission or
227 pathogenicity although pending investigation will provide confirmatory results on its
228 significance.

229

230 **DISCUSSION:**

231 We have developed a novel method which combines specific depletion and
232 enrichment strategies that results in efficient SARS-CoV-2 RNA-seq with high genome
233 coverage and depth. An advantage of this protocol is that it generates sequence data

234 directly from swab specimens without the need to passage the virus in cell culture thereby
235 reducing the handling of infectious material and induction of culture-acquired mutations.
236 Another obstacle in sequencing directly from swab specimens is that most FDA-approved
237 commercially available RNA extraction kits are specifically optimized to recover low
238 amounts of total nucleic acids, include carrier polyA RNA that could be convertible into
239 sequence able molecules, as has been observed previously with RNA-seq of Lassa- or
240 Ebola-positive clinical specimens (17).

241 The workflow incorporates amplification of low-abundance RNA into micrograms
242 of DNA, followed by conversion from a fraction of the DNA into Illumina-compatible
243 sequencing libraries and enrichment of these libraries for SARS-CoV-2 sequences. In
244 addition, during the reverse transcription step a reagent was incorporated to reduce the
245 subsequent amplification of host ribosomal RNA. This approach is robust in that it
246 converts low amounts of RNA into microgram quantities of DNA representative of all the
247 RNA species (aside of rRNA) present in the specimen. This DNA can be stored
248 indefinitely to be interrogated by multiple techniques at a later date. Additionally, RNA
249 amplification is likely less sensitive to low viral abundance compared to RT-PCR. Finally,
250 the use of probes to enrich for coronavirus-specific sequencing library molecules is less
251 sensitive to variants compared to tiling PCR amplicon approaches (18-22).

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

257 put in place, there was less mobility of people within and between states. It is possible
258 that these measures, compounded by potential inherent transmission variability of some
259 viral isolates, influenced the change in the frequency of D614G, clades and P323L/F that
260 we noted during this time period within Nevada. In addition, we also found 379C>A with
261 a high prevalence in our study specimens compared to the subsampling of sequences
262 from the United States and globally (Figure 6). This is a synonymous mutation in nsp1,
263 hence the biological relevance of this nucleotide variant remains to be elucidated.

264 Of the 14,885 complete SARS-CoV-2 genomes currently available (as of August
265 14, 2020) in NCBI there are only 6 genomes that have the P323F variant (accession
266 number: MT706208, LR860619, MT345877, MT627429, MT810889, MT811171). In this
267 study 62 of the 133 specimens from Northern Nevada contain P323F. That is 46% of
268 specimens from Northern Nevada contain P323F compared to 0.04% of NCBI deposited
269 SARS-CoV-2 isolates. We performed structural modeling of the P323L/F variation and
270 did not find any significant change to the nsp12 conformation with either P323L or P323F,
271 therefore this variant is most likely a neutral mutation and does not confer either a fitness
272 advantage or disadvantage to transmission or pathogenicity of SARS-CoV-2.

273 We find that the overall trend for D614G in Nevada during this time period was
274 similar to what was observed in other states and internationally, with the exception of
275 within Asia where the D614 allele had originated. We noted that there were differences
276 between Northern Nevada and Southern Nevada. In Northern Nevada clade 20C and
277 F323 were more frequent, while during this same time period in Southern Nevada clade
278 20A and L323 were more prevalent. These data indicate that there were distinct genomic
279 profiles of the SARS-CoV-2 viruses that were circulating in these populations during the

280 initial months of the pandemic while stay-at-home order were in place to help prevent
281 transmission of the virus.
282
283 **MATERIALS AND METHODS:**
284 **SARS-CoV-2 specimen and library preparation.**
285 Nasal and Nasopharyngeal swab specimens were received at the Nevada State
286 Public Health Lab (NSPHL) or Southern Nevada Public Health Lab (SNPHL) and RNA
287 extraction was completed using either a QIAamp Viral RNA Mini Kit (QIAGEN) or Mag-
288 Bind Viral DNA/RNA kit (Omega Biotek). Specimens were tested for the presence of
289 coronaviral RNA using FDA-approved kits that employed RT-PCR to detect SARS-COV-
290 2 RNA.
291 A set of 200 coronavirus positive specimens were selected for genome
292 sequencing. Specimens were treated with DNase I (QIAGEN) for 30 minutes at room
293 temperature and concentrated using RNeasy Minelute spin columns (QIAGEN) based on
294 the manufacturer supplied protocol. These concentrated samples were converted into
295 Illumina-compatible sequencing libraries with a QIAseq FX Single Cell RNA Library kit
296 (QIAGEN). RNA samples were annealed to a 1:12.5 dilution of QIAseq FastSelect -HMR
297 probes (QIAGEN) to reduce subsequent amplification of human ribosomal RNA. After
298 treatment to remove trace DNA from the samples, a reverse transcription reaction was
299 carried out using random hexamers. The synthesized DNA was ligated to one another,
300 followed by isothermal linear amplification. Amplified DNA (1 μg) was enzymatically
301 sheared to an average insert size of 300 bp, and Illumina-compatible dual-indexed
302 sequencing adapters were ligated to the ends. Next, about 300 ng of adapter-ligated

303 sample was amplified with 6 cycles of PCR with KAPA HiFi HotStart polymerase (Roche
304 Sequencing Solutions). Enrichment of library molecules containing SARS-CoV-2
305 sequence was conducted with a myBaits kit and coronavirus-specific biotinylated probes
306 (Arbor Biosciences). Each enrichment used 500 ng of PCR-amplified DNA, was carried
307 out based on manufacturer instructions at a hybridization temperature of 65° for 16 hours,
308 and was completed with 8-16 cycles of PCR using KAPA HiFi HotStart polymerase.
309 Samples were sequenced using an Illumina Next-seq mid-output (2 x 75). The generated
310 FASTQ files from the sequencing reaction were analyzed as described below. The data
311 files are available at GISAID, NCBI under the
312 https://www.ncbi.nlm.nih.gov/bioproject/657893

313 **Computational analysis.**

314 Sequence pair libraries were trimmed using Trimmomatic, version 0.39 and
315 adapter-clipping setting "2:30:10:2:keepBothReads" (23). Read pairs were aligned
316 against the Wuhan reference genome (NC_045512.2) by Bowtie 2, version 2.3.5, local
317 alignment (24). PCR optical duplicates were removed via Picard MarkDuplicates (25).

318 Variants were called using Freebayes, version 1.0.2, with ploidy set to 1, minimum
319 allele frequency 0.75, and minimum depth of 4 (26). No variants were called in the first
320 200 bp and final 63 bp of the COVID-19 genome. High-quality variant sites were selected
321 where site "QUAL > 20" using *vcffilter*, VCFlib version 1.0.0_rc2 (27). Individual genomes
322 were reconstructed by their filter-passing variants using *bcftools consensus* and only
323 where aligned coverage depth ≥ 4; bases with coverage below four are reported as
324 unknown (Ns) (28).

325 A set of 3,644 complete, high-coverage SARS-CoV-2 genomes reported in the
326 July 15, 2020 Nextstrain.org global subsample and metadata were obtained from GISAID
327 and combined with our own samples to determine their phylogenetic placement (29, 30).
328 Four of the global samples were set aside after screening for unexpected FASTA
329 characters. The combined sets of global and Nevadan samples were aligned together,
330 with metadata, by the *augur* phylodynamic pipelines of the *ncov* build of the *nextstrain*
331 command-line tool, version 2.0.0.post1 (31).

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

335 **nsp12 protein modeling.**

336 Sequence of nsp12 (RdRp) protein for SARS-CoV-2 (YP_009725307.0) was
337 retrieved from NCBI protein database and 3D model was structured based on a previously
338 published report (PDB ID: 6XEZ (32). In addition to nsp12 (chain A), the model also
339 contains nsp7 (chain C), nsp8 (chain B and D), nsp13 (chain E and F), ligands (Zn^{2+},
340 Mg^{2+}) and RNA template and product strands. Mutational changes to residue 323 within
341 nsp12 were performed using PyMol Molecular Graphics System (version 2.0, Schrödinger
342 LLC). The original proline (P) was mutated to either leucine (L) or phenylalanine (F) as
343 indicated, these residues along with residues containing side chains within 5 Å of P323L/F
344 are shown as sticks. The rotamers for each P323L/F were assessed and those with the
345 least rotational strain and steric hindrance were used to generate the final image. To
346 determine any NCBI deposited sequences which contain the P323F variant, standard
347 protein BLAST from the BLASTp suite was used to find nsp12 protein sequences which

348 contained FSTVFP**F**TSFGP (P323F is bold and underlined) from full length SARS-CoV-
349 2 genomes. The P323F amino acid changes were confirmed with the NCBI deposited
350 nucleotide sequences.
351
352 **DATA AVAILABILITY:**
353 All sequences are available at bio project:
354 https://www.ncbi.nlm.nih.gov/bioproject/657893. All reported data are deposited and
355 available at GISAID with the following accession numbers: EPI_ISL_515409,
356 EPI_ISL_515407, EPI_ISL_515408, EPI_ISL_515375, EPI_ISL_515376,
357 EPI_ISL_515373, EPI_ISL_515374, EPI_ISL_515371, EPI_ISL_515372,
358 EPI_ISL_515370, EPI_ISL_515416, EPI_ISL_515417, EPI_ISL_515414,
359 EPI_ISL_515415, EPI_ISL_515379, EPI_ISL_515412, EPI_ISL_515413,
360 EPI_ISL_515377, EPI_ISL_515410, EPI_ISL_515378, EPI_ISL_515411,
361 EPI_ISL_515364, EPI_ISL_515365, EPI_ISL_515362, EPI_ISL_515363,
362 EPI_ISL_515360, EPI_ISL_515361, EPI_ISL_515405, EPI_ISL_515406,
363 EPI_ISL_515403, EPI_ISL_515404, EPI_ISL_515368, EPI_ISL_515401,
364 EPI_ISL_515369, EPI_ISL_515402, EPI_ISL_515366, EPI_ISL_515367,
365 EPI_ISL_515400, EPI_ISL_515353, EPI_ISL_515354, EPI_ISL_515351,
366 EPI_ISL_515352, EPI_ISL_515350, EPI_ISL_515359, EPI_ISL_515357,
367 EPI_ISL_515358, EPI_ISL_515355, EPI_ISL_515356, EPI_ISL_515342,
368 EPI_ISL_515343, EPI_ISL_515340, EPI_ISL_515461, EPI_ISL_515341,
369 EPI_ISL_515462, EPI_ISL_515460, EPI_ISL_515348, EPI_ISL_515349,
370 EPI_ISL_515346, EPI_ISL_515347, EPI_ISL_515344, EPI_ISL_515345,

371 EPI_ISL_515331, EPI_ISL_515298, EPI_ISL_515452, EPI_ISL_515332,
372 EPI_ISL_515299, EPI_ISL_515453, EPI_ISL_515296, EPI_ISL_515450,
373 EPI_ISL_515330, EPI_ISL_515297, EPI_ISL_515451, EPI_ISL_515294,
374 EPI_ISL_515295, EPI_ISL_515293, EPI_ISL_515339, EPI_ISL_515337,
375 EPI_ISL_515458, EPI_ISL_515338, EPI_ISL_515459, EPI_ISL_515335,
376 EPI_ISL_515456, EPI_ISL_515336, EPI_ISL_515457, EPI_ISL_515333,
377 EPI_ISL_515454, EPI_ISL_515334, EPI_ISL_515455, EPI_ISL_515319,
378 EPI_ISL_515320, EPI_ISL_515441, EPI_ISL_515321, EPI_ISL_515442,
379 EPI_ISL_515440, EPI_ISL_515328, EPI_ISL_515449, EPI_ISL_515329,
380 EPI_ISL_515326, EPI_ISL_515447, EPI_ISL_515327, EPI_ISL_515448,
381 EPI_ISL_515324, EPI_ISL_515445, EPI_ISL_515325, EPI_ISL_515446,
382 EPI_ISL_515322, EPI_ISL_515443, EPI_ISL_515323, EPI_ISL_515444,
383 EPI_ISL_515308, EPI_ISL_515429, EPI_ISL_515309, EPI_ISL_515390,
384 EPI_ISL_515397, EPI_ISL_515430, EPI_ISL_515310, EPI_ISL_515398,
385 EPI_ISL_515431, EPI_ISL_515395, EPI_ISL_515396, EPI_ISL_515393,
386 EPI_ISL_515394, EPI_ISL_515391, EPI_ISL_515392, EPI_ISL_515317,
387 EPI_ISL_515438, EPI_ISL_515318, EPI_ISL_515439, EPI_ISL_515315,
388 EPI_ISL_515436, EPI_ISL_515316, EPI_ISL_515437, EPI_ISL_515313,
389 EPI_ISL_515434, EPI_ISL_515314, EPI_ISL_515435, EPI_ISL_515311,
390 EPI_ISL_515399, EPI_ISL_515432, EPI_ISL_515312, EPI_ISL_515433,
391 EPI_ISL_515418, EPI_ISL_515419, EPI_ISL_515386, EPI_ISL_515387,
392 EPI_ISL_515420, EPI_ISL_515384, EPI_ISL_515385, EPI_ISL_515382,
393 EPI_ISL_515383, EPI_ISL_515380, EPI_ISL_515381, EPI_ISL_515306,

394 EPI_ISL_515427, EPI_ISL_515307, EPI_ISL_515428, EPI_ISL_515304,
395 EPI_ISL_515425, EPI_ISL_515305, EPI_ISL_515426, EPI_ISL_515302,
396 EPI_ISL_515423, EPI_ISL_515303, EPI_ISL_515424, EPI_ISL_515300,
397 EPI_ISL_515388, EPI_ISL_515421, EPI_ISL_515301, EPI_ISL_515389,
398 EPI_ISL_515422
399

400 **ACKNOWLEDGEMENTS:**

401 We thank the staff at the Nevada State Public Health Lab (NSPHL) and Southern
402 Nevada, Public Health Laboratory (SNPHL) for providing the RNA extracts from the
403 patient's specimen. This work was supported by University of Nevada, Reno Vice
404 President for Research and Innovation (VPRI), Department of Microbiology &
405 Immunology, UNR School of Medicine, Nevada IDeA Network of Biomedical Research
406 Excellence (INBRE) from the National Institute of General Medical Sciences (GM 103440
407 and GM 104944) from the National Institutes of Health (NIH). The authors wish to
408 acknowledge the support of Research & Innovation and the Office of Information
409 Technology at the University of Nevada, Reno for computing time on the Pronghorn High-
410 Performance Computing Cluster.
411

412 **AUTHOR CONTRIBUTION:**

413 PDH: conceptualization, formal analysis, methodology, writing – original draft preparation,
414 writing – review and editing
415 RLT: conceptualization, formal analysis, methodology, writing – original draft preparation,
416 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: 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. 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 5:536-544.
2. Petrosillo N, Viceconte G, Ergonul O, Ippolito G, Petersen E. 2020. COVID-19, SARS and MERS: are they closely related? Clin Microbiol Infect 26:729-734.
3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579:270-273.
4. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, Yang L, Ding C, Zhu X, Lv R, Zhu J, Hassan B, Feng Y, Tan W, Wang C. 2018. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect 7:154.
5. de Wit E, van Doremalen N, Falzarano D, Munster VJ. 2016. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 14:523-34.
6. Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT, Cheng VC, Chan KH, Tsang DN, Yung RW, Ng TK, Yuen KY, group Ss. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361:1319-25.
7. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367:1814-20.
8. Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, Chan KH, Li PH, Tan SY, Chang Q, Xie JP, Liu XQ, Xu J, Li DX, Yuen KY, Peiris, Guan Y. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 362:1353-8.

469 9. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. 2020. The Architecture of
470 SARS-CoV-2 Transcriptome. Cell 181:914-921 e10.
471 10. Licastro D, Rajasekharan S, Dal Monego S, Segat L, D'Agaro P, Marcello A. 2020.
472 Isolation and Full-Length Genome Characterization of SARS-CoV-2 from COVID-
473 19 Cases in Northern Italy. J Virol 94.
474 11. Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, Halden RU,
475 Hogue BG, Scotch M, Lim ES. 2020. An 81-Nucleotide Deletion in SARS-CoV-2
476 ORF7a Identified from Sentinel Surveillance in Arizona (January to March 2020).
477 J Virol 94.
478 12. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, Hu Y, Tao ZW, Tian JH, Pei
479 YY, Yuan ML, Zhang YL, Dai FH, Liu Y, Wang QM, Zheng JJ, Xu L, Holmes EC,
480 Zhang YZ. 2020. A new coronavirus associated with human respiratory disease in
481 China. Nature 579:265-269.
482 13. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ,
483 Zambon M, Ellis J, Lewis PA, Hiscox JA, Matthews DA. 2020. Characterisation of
484 the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced
485 in-frame deletion of the furin-like cleavage site from the spike glycoprotein.
486 Genome Med 12:68.
487 14. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure,
488 Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181:281-
489 292 e6.
490 15. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner
491 N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG,
492 Evans CM, Freeman TM, de Silva TI, Sheffield C-GG, McDanal C, Perez LG, Tang
493 H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC. 2020.
494 Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases
495 Infectivity of the COVID-19 Virus. Cell 182:812-827 e19.
496 16. Nextclade. 2020. [cited 17 Aug 2020]. Available: https://cladesnextstrainorg.
497 17. Matranga CB, Andersen KG, Winnicki S, Busby M, Gladden AD, Tewhey R,
498 Stremlau M, Berlin A, Gire SK, England E, Moses LM, Mikkelsen TS, Odia I,
499 Ehiane PE, Folarin O, Goba A, Kahn SH, Grant DS, Honko A, Hensley L, Happi

500 C, Garry RF, Malboeuf CM, Birren BW, Gnirke A, Levin JZ, Sabeti PC. 2014.
501 Enhanced methods for unbiased deep sequencing of Lassa and Ebola RNA
502 viruses from clinical and biological samples. Genome Biol 15:519.
503 18. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, Lipkin WI. 2015.
504 Virome Capture Sequencing Enables Sensitive Viral Diagnosis and
505 Comprehensive Virome Analysis. mBio 6:e01491-15.
506 19. O'Flaherty BM, Li Y, Tao Y, Paden CR, Queen K, Zhang J, Dinwiddie DL, Gross
507 SM, Schroth GP, Tong S. 2018. Comprehensive viral enrichment enables sensitive
508 respiratory virus genomic identification and analysis by next generation
509 sequencing. Genome Res 28:869-877.
510 20. Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, Tong S. 2020. Rapid,
511 Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome
512 Coronavirus 2. Emerg Infect Dis 26.
513 21. Paskey AC, Frey KG, Schroth G, Gross S, Hamilton T, Bishop-Lilly KA. 2019.
514 Enrichment post-library preparation enhances the sensitivity of high-throughput
515 sequencing-based detection and characterization of viruses from complex
516 samples. BMC Genomics 20:155.
517 22. Xiao M, Liu X, Ji J, Li M, Li J, Yang L, Sun W, Ren P, Yang G, Zhao J, Liang T,
518 Ren H, Chen T, Zhong H, Song W, Wang Y, Deng Z, Zhao Y, Ou Z, Wang D, Cai
519 J, Cheng X, Feng T, Wu H, Gong Y, Yang H, Wang J, Xu X, Zhu S, Chen F, Zhang
520 Y, Chen W, Li Y, Li J. 2020. Multiple approaches for massively parallel sequencing
521 of SARS-CoV-2 genomes directly from clinical samples. Genome Med 12:57.
522 23. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
523 sequence data. Bioinformatics 30:2114-20.
524 24. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat
525 Methods 9:357-9.
526 25. Picard-Toolkit. 2019. [cited 17 Aug 2020]. In: Github [Internet] Available:
527 http://broadinstitute.github.io/picard.
528 26. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read
529 sequencing. arXiv [q-bioGN] Available: http://arxiv.org/abs/1207.3907.

530 27. Garrison E. 2019. Vcflib: A C++ library for parsing and manipulating VCF files.
531 Available: https://githubcom/vcflib/vcflib.
532 28. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association
533 mapping and population genetical parameter estimation from sequencing data.
534 Bioinformatics 27:2987-93.
535 29. Elbe S, Buckland-Merrett G. 2017. Data, disease and diplomacy: GISAID's
536 innovative contribution to global health. Glob Chall 1:33-46.
537 30. Nextstrain. 2020. Genomic epidemiology of novel coronavirus - Global
538 subsampling. In: Nextstrainorg [Internet] 15 Jul 2020 [cited 15 Jul 2020] Available:
539 https://nextstrain.org/ncov/global/2020-07-15?d=tree&l=clock&legend=closed.
540 31. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P,
541 Bedford T, Neher RA. 2018. Nextstrain: real-time tracking of pathogen evolution.
542 Bioinformatics 34:4121-4123.
543 32. Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi
544 K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. 2020.
545 Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2
546 Replication-Transcription Complex. Cell doi:10.1016/j.cell.2020.07.033.
547

548 **FIGURE LEGENDS:**

549 **Figure 1. Workflow of SARS-CoV-2 genome sequencing and analysis from**
550 **nasopharyngeal patient specimens in Nevada.** (A) RNA was extracted from Nasal or
551 Nasopharygeal (NP) swabs taken from patients in Nevada and first used to determine the
552 presence of SARS-CoV-2 genomes by RT-qPCR. Next generation sequencing (NGS)
553 libraries were prepared from positive specimens, this included steps for ribosomal RNA
554 depletion and SARS-CoV-2 enrichment. Subsequent libraries were pooled and used for
555 whole genome sequencing at the Nevada Genomics Center on the Illumina NextSeq 500
556 instrument. FASTQ files were aligned to the reference genome, and analyzed to
557 determine nucleotide variation and phylogenetic relationship. (B) A total of 200 specimens

558 were sequenced, of which 174 had over 99% coverage of the SARS-CoV-2 genome.
559 This included 133 patient specimens from Northern Nevada, 40 from Southern Nevada
560 and 1 specimen that was re-sequenced. (C) Correlation between RT-qPCR Ct value and
561 the percentage of coverage in the whole genome sequencing after trimming and
562 alignment. (D) Nucleotide variants across the SARS-CoV-2 genome in the 173 specimens
563 from Nevada from March 6 to June 5.
564
565 **Figure 2. Distribution of D614G in Nevada and comparison with the United States**
566 **and global proportion.** (A) Cumulative frequency of D614G in 173 patient specimens
567 from Nevada from March 6 to June 5, 2020 (D614 is indicated by teal, G614 is indicated
568 by yellow). Pie charts depict the cumulative proportion up to the indicated time point
569 (March 15, April 1, May 1, June 5). The total number of specimens included at each time
570 point is specified below each pie chart. Effective dates of emergency orders and
571 regulatory responses to SARS-CoV-2 spread in Nevada are indicated on the frequency
572 graph time axis. (B) Proportion of D614G in the United States from March 6 to June 5,
573 specimens from Nevada are divided in the geographic area that they originated from,
574 Northern Nevada (N-NV) includes 133 specimens from Washoe County, Carson-Tahoe,
575 and other northern counties, and Southern Nevada (S-NV) includes 40 specimens from
576 Clark County. (C) Global proportion of D614G in the shown regions during the same time
577 period from a subsampling of sequences deposited in Nextstrain.org. The size of the pie
578 chart corresponds to the relative specimen number for each region.
579

580 **Figure 3. Distribution of SARS-CoV-2 clades in Nevada.** (A) Cumulative frequency of
581 SARS-CoV-2 clades in 173 patient specimens from Nevada during March 6 to June 5.
582 The five clades are colored 19A (blue), 19B (teal), 20A (green), 20B (yellow) and 20C
583 (orange). Pie charts depict the cumulative proportion up to the indicated time point (March
584 15, April 1, May 1, June 5). The total number of specimens included at each time point is
585 specified below each pie chart. Dates of emergency orders and regulations meant to slow
586 the spread of SARS-CoV-2 in Nevada are indicated on the time scale of the frequency
587 graph. (B) Circular dendrogram depicting clades from Nevada specimens. (C) Pie chart
588 of the clades from northern Nevada (N-NV), southern Nevada (S-NV) and the United
589 States. (D) Pie charts show the proportion of clades from global regions during the same
590 time period from a subsampling of sequences deposited in Nextstrain.org. The size of the
591 pie chart corresponds to the relative specimen number for each region.
592
593 **Figure 4. Distribution of P323L/F (nsp12, RdRp) in Nevada.** (A) Cumulative frequency
594 of P323L/F (nsp12, RdRp) in 173 patient specimens from Nevada during March 6 to June
595 5. The amino acid at position 323 is indicated by teal for proline (P), yellow for leucine (L)
596 and blue for phenylalanine (F). Pie charts depict the cumulative proportion up to the
597 indicated time point (March 15, April 1, May 1, June 5). The total number of specimens
598 included at each time point is specified below each pie chart. Dates of emergency orders
599 and regulations meant to slow the spread of SARS-CoV-2 in Nevada are indicated on the
600 time scale of the frequency graph. (B) Circular dendrogram representing the distribution
601 of amino acid change at residue 323 of nsp12 from a global subsampling of sequences
602 deposited in Nextstrain.org from March 6 to June 5, the larger dots indicate specimens

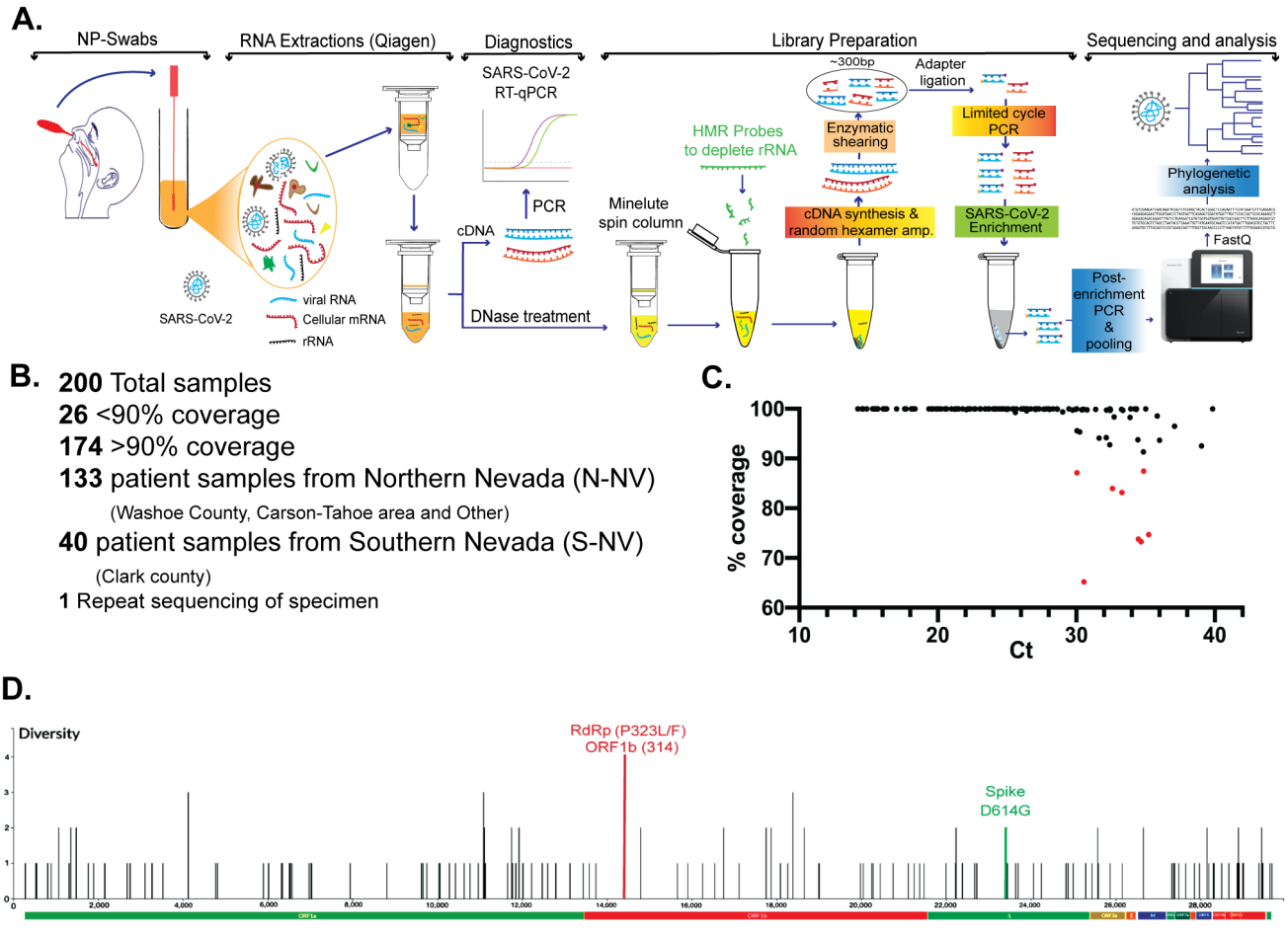
603 from Nevada. (C) Pie chart indicating the ratio of P/L/F in Northern NV and Southern NV
604 specimens from this study. (D) Proportion of P323L/F from a subsampling of sequences
605 deposited in Nextstrain.org for the United States and (E) global regions from March 6 to
606 June 5. The size of the pie chart corresponds to the relative specimen number for each
607 region.
608
609 **Figure 5. Structure of SARS-CoV-2 nsp12 (RdRp) P323L/F.** (a) Diagram depicting
610 ORF1b genomic location and encoded proteins. Below the linear protein schematic of
611 nsp12 specific nucleotide variants at position 14,407 and 14,408 and the resulting amino
612 acid changes are indicated. (b) SARS-CoV-2 replicase complex modeled from 6XEZ
613 template. This model includes nsp12, nsp7, nsp8, nsp13, ligands (ZN, Mg), the template
614 and product strand of RNA. P323L/F within the interface domain of nsp12 is located at
615 the left side of the complex. (c) Model nsp12 showing P323L/F within the interface
616 domain. Residue 323 is shown with either P, L or F and amino acids with side chains
617 within 5 Å of residue 323 are depicted as sticks in the cartoon model.
618
619 **Figure 6. Distribution of nucleotide variant 379C>A.** (a) Green line at the far-left end
620 of the genome denoted nucleotide position 379 of nsp1. (b) Circular dendrogram of global
621 subsample of sequences from Nextstrain.org with NV specimens indicated by larger dots.
622 (c) Pie chart indicating the proportion of sequences with either the cytidine (C) or
623 adenosine (A) at position 379 from the Nevada specimens. Subsample of sequences from
624 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.

648 **Figure 1:**

649

650



651

652

653

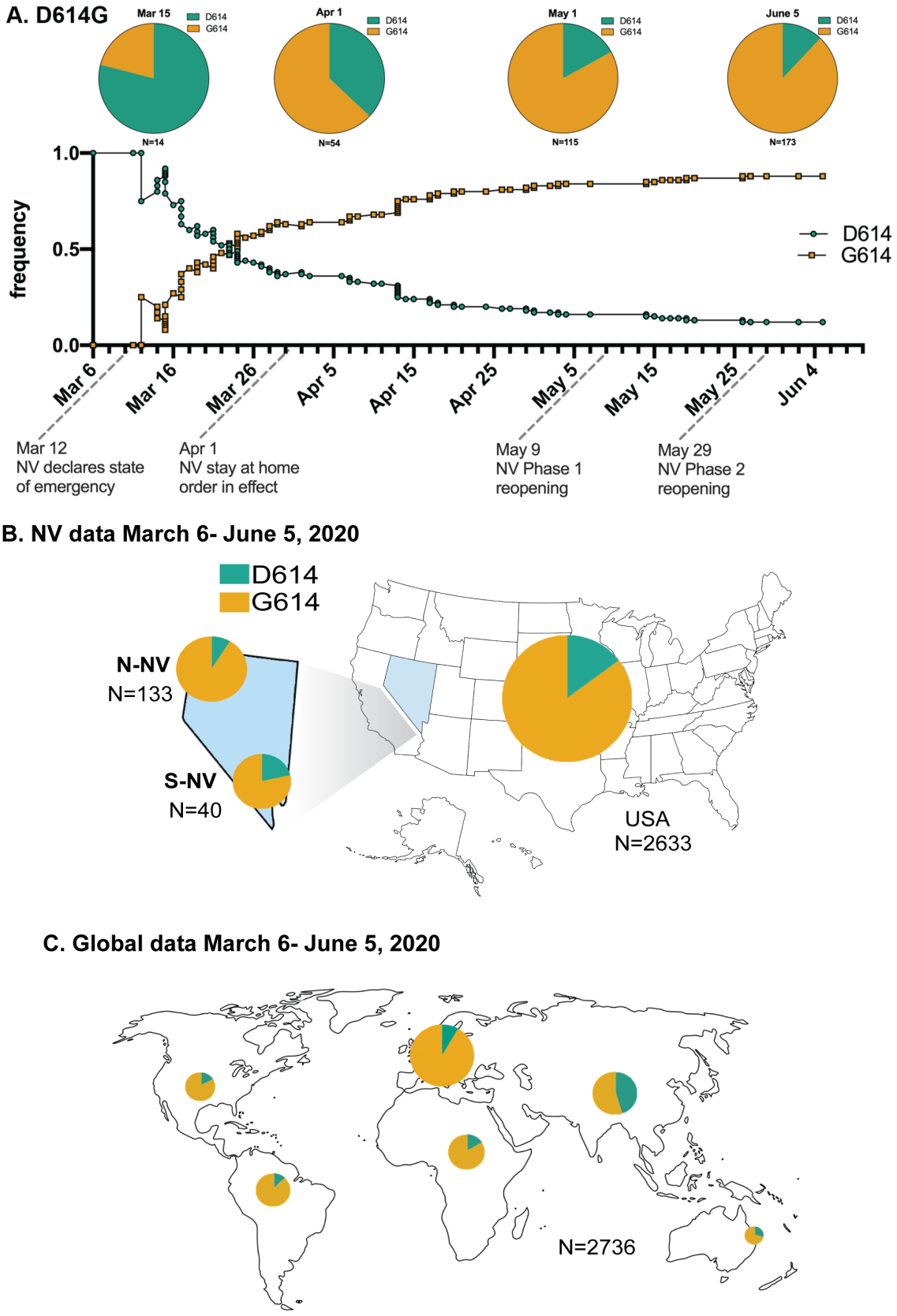
654

655

656

657

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

681 **Figure 3:**

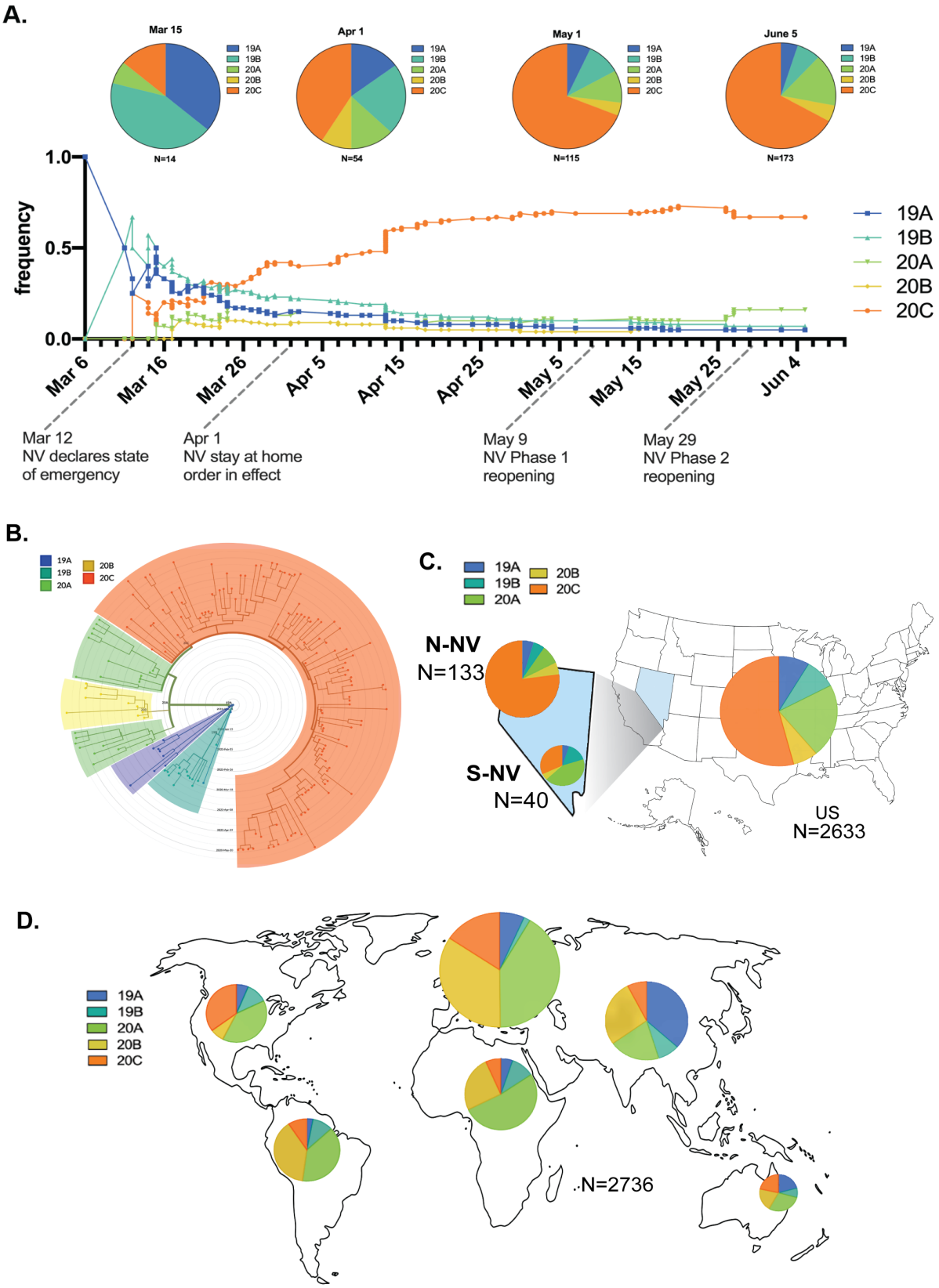

A.
Mar 15
Apr 1
May 1
June 5
19A
19B
20A
20B
20C
N=14
N=54
N=115
N=173
1.0
0.5
0.0
frequency
Mar 6
Mar 16
Mar 26
Apr 5
Apr 15
Apr 25
May 5
May 15
May 25
Jun 4
Mar 12
NV declares state of emergency
Apr 1
NV stay at home order in effect
May 9
NV Phase 1 reopening
May 29
NV Phase 2 reopening
B.
C.
N-NV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

Figure 4:

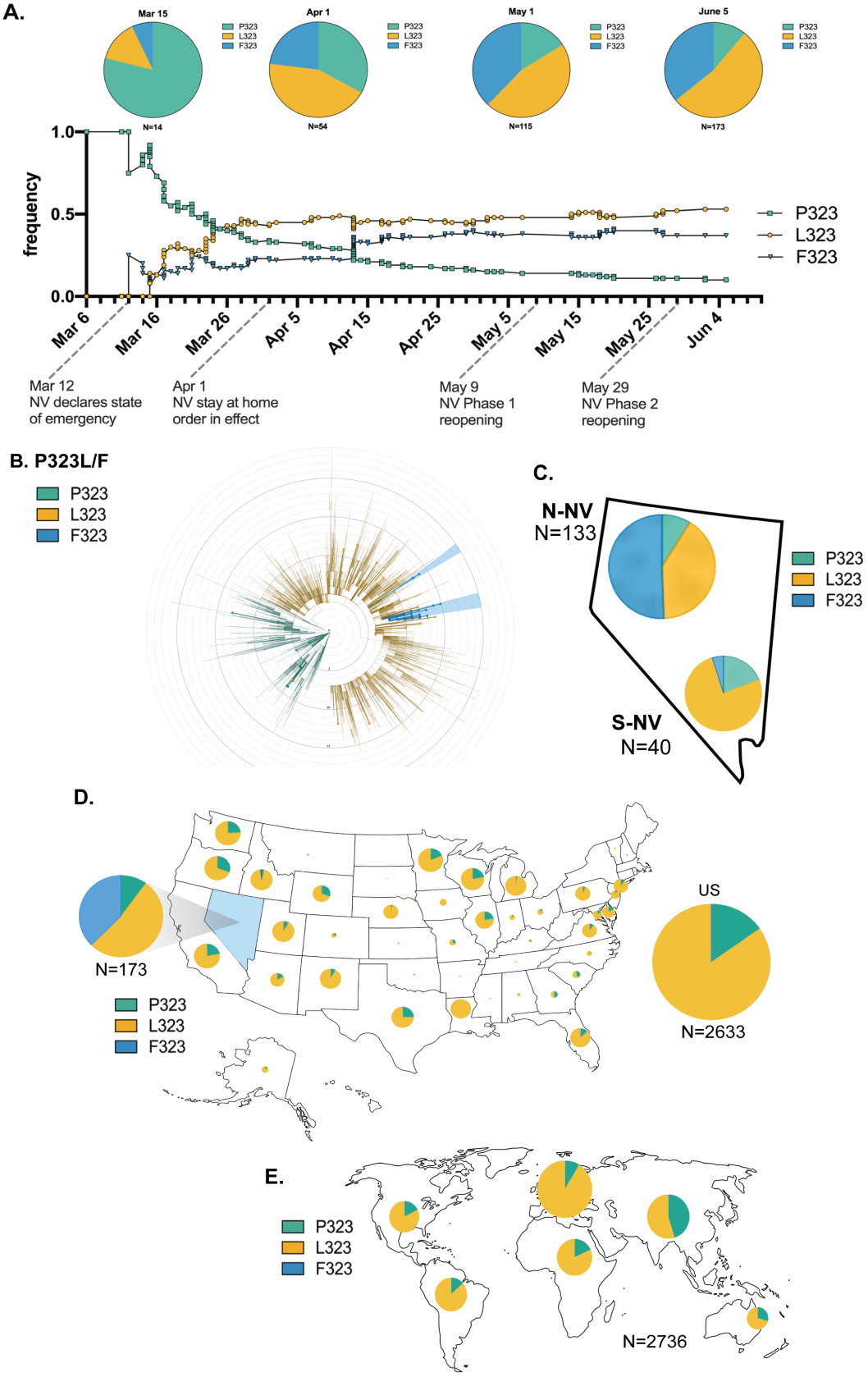

A.
Mar 15
Apr 1
May 1
June 5
P323
L323
F323
N=14
N=54
N=115
N=173
frequency
1.0
0.5
0.0
Mar 6
Mar 16
Mar 26
Apr 5
Apr 15
Apr 25
May 5
May 15
May 25
Jun 4
Mar 12
NV declares state of emergency
Apr 1
NV stay at home order in effect
May 9
NV Phase 1 reopening
May 29
NV Phase 2 reopening
B. P323L/F
C.
N-NV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

Figure 5:

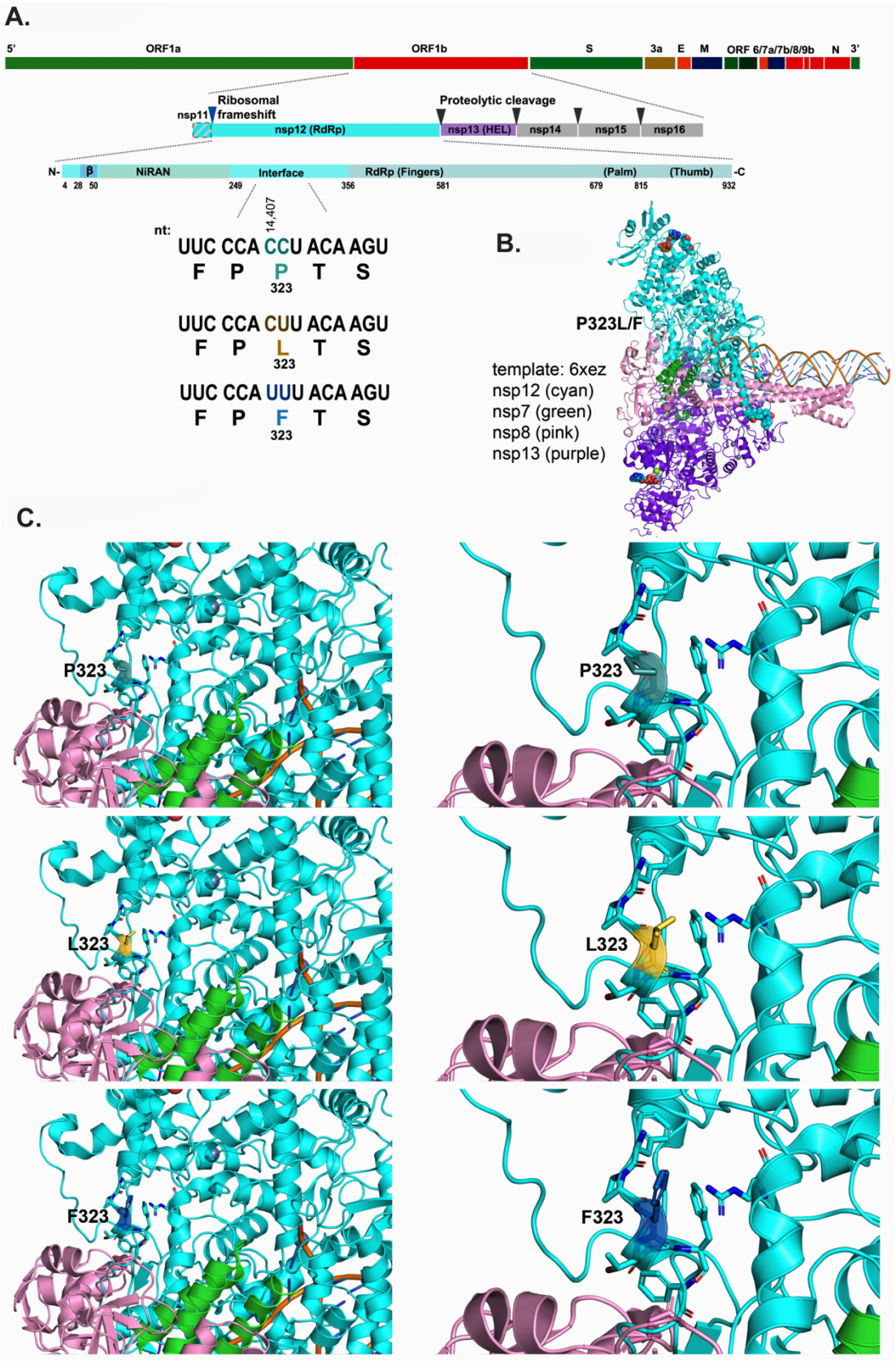
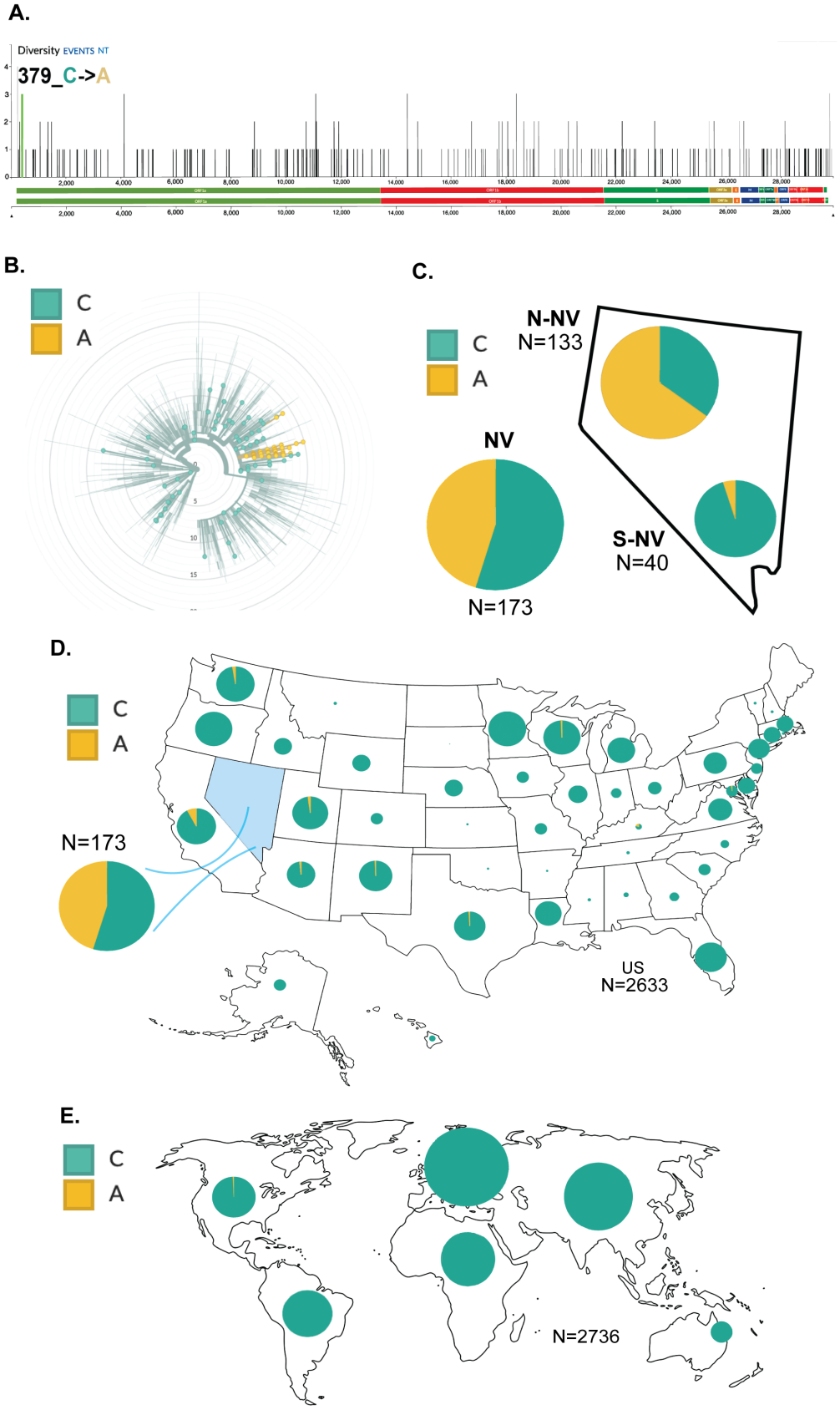

A.
5'
ORF1a
ORF1b
S
3a
E
M
ORF 6/7a/7b/8/9b
N
3'
Ribosomal frameshift
nsp11
nsp12 (RdRp)
Proteolytic cleavage
nsp13 (HEL)
nsp14
nsp15
nsp16
N-
β
NiRAN
Interface
RdRp (Fingers)
(Palm)
(Thumb)
-C
4 28 50
249
356
581
679
815
932
14,407
nt:
UUC CCA CCU ACA AGU
F P P T S
323
UUC CCA CUU ACA AGU
F P L T S
323
UUC CCA UUU ACA AGU
F P F T S
323
B.
P323L/F
template: 6xez
nsp12 (cyan)
nsp7 (green)
nsp8 (pink)
nsp13 (purple)
C.
P323
P323
L323
L323
F323
F323

Figure 6:


A.
Diversity EVENTS NT
379_C->A
B.
C
A
C.
C
A
N-NV
N=133
NV
N=173
S-NV
N=40
D.
C
A
N=173
US
N=2633
E.
C
A
N=2736