Full vaccination against COVID-19 suppresses SARS-CoV-2 delta variant and spike gene mutation frequencies and generates purifying selection pressure

Ting-Yu Yeh and Gregory P. Contreras

Agricultural Biotechnology Laboratory, Auxergen Inc., Columbus Center, 701 E Pratt Street, Baltimore, MD 21202, USA

*Ting-Yu Yeh is Corresponding author

E-mail: yehty@auxergen.com

Tel: 1-443-762-1974

Key words: COVID-19, SARS-CoV-2, mutation frequency, neutrality tests, polymorphism, site-frequency spectrum

Short running title: Purifying selection of SARS-CoV-2 by vaccination

Word count: 2329 words (text)

This paper is dedicated to Professor Austin Hughes (1949-2015), an evolution biologist at University of South Carolina, for his lifetime achievement in neural theory and adaptive molecular evolution.

Abstract

COVID-19 vaccination resistance has become a major challenge to prevent global SARS-CoV-2 transmission. Here we report that the vaccination coverage rate is inversely correlated to the mutation frequency of the full genome (R^2=0.878) and spike gene (R^2=0.829) of SARS-CoV-2 delta variants in 16 countries, suggesting that full vaccination against COVID-19, with other mitigation strategies, is critical to suppress emergent mutations. Neutrality analysis of *DH* and Zeng's *E* tests suggested that directional selection was the major driving force of delta variant evolution. To eliminate the homogenous effects (population expansion, selective sweep etc.), the synonymous (D_{syn}) and nonsynonymous (D_{nonsyn}) polymorphisms of the delta variant spike gene were estimated with Tajima's *D* statistic. Both *D* ratio (D_{nonsyn}/D_{syn}) and ΔD (D_{syn}-D_{nonsyn}) have positive correlation with the full vaccination rate (R^2= 0.723 and 0.505, respectively) in 19 countries, indicating that purifying selection pressure of SARS-CoV-2 spike gene increased as the vaccination coverage rate increased. Taken together, our data suggests that vaccination plays an important role in the purifying selection force of spike protein of SARS-CoV-2 delta variants.

45 **Introduction**

46 Legally required vaccination against various infectious diseases is essential to public
47 health policy in many countries. However, resistance to voluntary COVID-19
48 vaccination has emerged worldwide, including in the United States. In mid-June
49 2021, 33% of Americans said they were not willing to be vaccinated (Kirzinger et al,
50 2021). Public distrust has undermined COVID-19 vaccine acceptance, in association
51 with a belief that the vaccine is ineffective, dangerous or compromises individual
52 freedom (Schmelz et al, 2021). Therefore, overcoming COVID-19 vaccination
53 resistance has become a major challenge to prevent global SARS-CoV-2
54 transmission.

55

56 Mutations drive genome variability, generating many different SARS-CoV-2 variants
57 as the virus evolves to escape vaccine-mediated immunity and thereby, develop
58 drug or vaccine resistance (Roy et al., 2020). The delta (lineage B1.617.2) variants
59 were first documented in October 2020 in India and were designated variants of
60 concern by the World Health Organization in May of 2021 due to its increased
61 transmissibility and virulence (https://www.who.int/en/activities/tracking-SARS-
62 CoV-2-variants/). All SARS-CoV-2 mutations result from two major mechanisms: (1)
63 spontaneous substitution/deletion of nucleotides, and (2) RNA recombination (Roy
64 et al, 2020, Yeh and Contreras, 2020, Ignatieva et al, 2021, Yeh and Contreras, 2021).
65 Until now, it remained unclear exactly how human vaccinations affected virus
66 selective pressure.

67

68

69 **Is there correlation between mutation frequency and vaccination?**

70 To explore this question, we first analyzed the correlation between the rates of full
71 vaccination and the point mutation frequency (*Mf*) of COVID-19 delta variants'
72 genome in 20 countries. Complete SARS-CoV-2 genome sequences with high
73 coverage from June 20 to July 3 2021 in 20 countries: Australia (*N*=121), France
74 (*N*=788), Germany (*N*=955), Indonesia (*N*=97), India (*N*=171), Ireland (*N*=617),
75 Israel (*N*=333), Italy (*N*=642), Japan (*N*=105), Mexico (*N*=368), Netherland (*N*=456),
76 Norway (*N*=142), Portugal (*N*=782), Singapore (*N*=131), Spain (*N*=689),
77 Switzerland (*N*=131), Sweden (*N*=786), Turkey (*N*=428), United States (*N*=537), and
78 United Kingdom (UK, *N*=3534) were collected from the Global Initiative on Sharing
79 All Influenza Data (GISAID) (https://www.gisaid.org/). *Mf* was calculated as Pi / (Ln
80 × Ns). Pi is the number of instances of polymorphism detected within the
81 genome/locus. Ln is the nucleotide length of the genome/locus. Ns is the number of
82 sequenced entities present in the dataset (Roy et al., 2020). The information of fully
83 vaccinated rate on 3 July 2021 in individual counties were obtained in Our World in
84 Data (https://ourworldindata.org/covid-vaccinations).

85

86 We found that *Mf* was logarithmically reduced as the full vaccination rate increased
87 in 16 of 20 countries (R^2=0.878, $P<10^{-7}$, Figure 1A). To our knowledge, this is the
88 first evidence suggesting that vaccinations could successfully suppress viral
89 mutations. Since the spike protein is the target of the vaccination program, we
90 further examined mutations of delta variant spike gene. Likewise, we found that

91 the *Mf* of the spike gene was also logarithmically reduced as the full vaccination rate
92 increased this time in 17 of the 20 countries (R^2=0.829, $P<10^{-6}$, Figure 1B). With a
93 10.8% vaccinated rate, *Mf* was exceptionally low in Australia, likely as a result of the
94 strict state lockdown restrictions. In contrast, the *Mfs* are higher in Japan,
95 Switzerland, and United States, suggesting that their social mitigation strategies
96 have been less successful. This also suggests that he mutation frequency may be not
97 only affected by vaccination only, but may also reflect overall control of the virus.
98

99 **How does vaccination affect selection pressure?**

100 SARS-CoV-2 is subject to many powerful selection pressures which may contribute
101 to vaccine escape. Like other viruses, the delta variant appears to have been selected
102 for its increased transmissibility regardless of vaccination. Previously we reported
103 on the mutations and selection pressure of SARS-CoV-2 among passengers in close-
104 quarters quarantine on the Diamond Princess cruise ship between January and
105 March 2020 (Yeh and Contreras, 2021). Until then, it had been unclear whether
106 COVID-19 variants were evolving randomly or under selection pressures generated
107 by vaccination and/or other mitigation efforts.
108

109 To understand how selection pressure operates within COVID-19 virus populations,
110 we analyzed the polymorphism of the full SARS-CoV-2 genome using neutrality tests
111 based on the site-frequency spectrum: (1) Tajima's *D* test, (2) normalized Wu and
112 Fay's *DH* test, and (3) Zeng's *E* test, using DNASP6 software with the bat coronavirus
113 RaTG13 as the outgroup sequence (Rozas, et al., 2017). The Tajima *D* test compares

114 pairwise nucleotide diversity (π, the average number of nucleotide differences per
115 site between two sequences) and total polymorphism to infer selection and/or
116 demographic events (Tajima, 1989). *DH* test is sensitive to the changes in high-
117 frequency variants and primarily affected by directional selection and no other
118 driving forces, such as demographic expansion (Zeng et al., 2006). Zeng's *E* test
119 contrasts high and low frequency regions of frequency spectrum, so this test can
120 detect population growth after a sweep (Zeng et al., 2006). Statistical significance of
121 observed values of all tests was obtained by coalescent simulation with free
122 recombination after 10,000 replicates. The probability for each test statistic was
123 calculated as the frequency of replicates with a value lower than the observed
124 statistic (two-tailed test) by DNASP6 (Rozas, et al., 2017).
125
126 We observed that Tajima's *D* values were negative and significantly deviated from
127 zero (-2.199 to -2.878, $P < 0.05$) among 19 countries, indicating an excess of variants
128 of low frequency (Table 1). Because the negative Tajima's *D* values could result from
129 demographic expansion and/or purifying/positive selection, we further employed
130 normalized *DH* (insensitive to population growth) and Zeng's *E* test (Zeng et al.,
131 2006). Because fixation causes quick loss of high-frequency alleles by drift, and
132 excess of low-frequency mutations are generated, Zeng's *E* test is very sensitive to
133 population growth. Genomic polymorphism of these viral populations revealed
134 significantly negative *DH* values (-1.177 to -2.791, $P<0.05$), but *E* values were not
135 significantly different from zero (-0.04 to 0.482, $P>0.05$) (Table 1). Similar results
136 were obtained after confining our analysis to the spike gene of SARS-CoV-2 delta

137 variants using Tajima' *D*, *DH*, and Zeng's *E* tests (Table 1). Taken together, these data
138 suggest that directional selection was operating in these countries, and
139 demographic expansion was less likely the major driving force.
140
141 Selection pressure has been observed in unvaccinated SARS-CoV-2 populations,
142 such as in the aforementioned Diamond Princess shipboard lockdown (Yeh and
143 Contreras, 2021). Our current results lead to the hypothesis that vaccination could
144 contribute to a part of the selection pressures on SARS-CoV-2, especially on the
145 spike protein. It has been shown that, within populations, the frequency distribution
146 of non-synonymous polymorphisms is negatively skewed relative to the distribution
147 of synonymous polymorphisms under purifying selection (Hahn et al, 2002; Hughes,
148 et al. 2005). Therefore, this hypothesis can be examined using a modified Tajima's *D*
149 statistics, which takes more negative values for non-synonymous (D_{nonsyn}) than for
150 synonymous sites (D_{syn}) of spike gene under purifying selection (Hahn et al, 2002,
151 Hughes, et al., 2005). The average number of pairwise synonymous differences (kS)
152 and the average number of nonsynonymous pairwise differences (kN), the number
153 of synonymous segregating sites (S_S), and the number of nonsynonymous
154 segregating sites (S_N) were computed with equation as Austin Hughes described (1)
155 $S^*_S=S_S/a1$, (2) $S^*_N=S_N/a1$, where a1 is as $a_1 = \sum_{i=1}^{n-1} \frac{1}{i}$. (Tajima, 1989, Hughes, et al.,
156 2005). D_{syn} was defined as kS − S^*_S, divided by the standard error of that difference,
157 and D_{nonsyn} was defined as kN − S^*_N, divided by the standard error of that difference
158 (Tajima, 1989, Hughes, et al., 2005). The values of K_a/K_s (the ratio of the number of
159 nonsynonymous substitutions per non-synonymous site (K_a), to the number of

160 synonymous substitutions per synonymous site (K_s), in a given period of time), and
161 the neutrality index of the McDonald-Kreitman (MK) test of the spike protein
162 (RaTG13 as the outgroup sequence) were analyzed by DNASP6 software (Rozas, et
163 al., 2017).
164
165 Unlike Tajima's *D* test, it has been demonstrated that D_{nonsyn} and D_{syn} analysis is
166 independent of sample size, so the value of D_{nonsyn} or D_{syn} can be compared between
167 data sets of different sizes (Hughes et al., 2008). The same excess of low-frequency
168 alleles in non-synonymous polymorphism was also observed (D_{nonsyn} = -1.482 to
169 −2.714; $P < 0.01$). In addition, the K_a/K_s values were less than 1 and the neutrality
170 index values of the MK test were significantly more than 1 ($P<0.05$) in all countries
171 (Table 1). These results confirmed that purifying selection led to constraint on the
172 available neutral mutations at non-synonymous sites of the spike gene of delta
173 variants, consistent to previous reports of SARS-CoV-2 variants in 2020 (Chaw et al,
174 2020).
175
176 Because non-synonymous and synonymous mutations are affected unequally by
177 purifying selection, D_{nonsyn} is expected to be disproportionally lower than D_{syn} (Hahn
178 et al., 2002). It has been proposed that the value of ΔD (D_{syn}-D_{nonsyn}) increased as
179 purifying selection became stronger, and ΔD (D_{syn}-D_{nonsyn}) is also able to rule out the
180 homogenous effects (population expansion, selective sweep etc.) in neutrality tests
181 (Hahn et al, 2002). We observed that ΔD of the spike gene of SARS-CoV-2 delta
182 variants is positively proportional to fully vaccination coverage rate (R^2= 0.505,

183 P=0.001). When using D ratio (D_{nonsyn}/D_{syn}), the standard error terms in both
184 equations of D_{nonsyn} and D_{syn} cancel out the sample size effect (Hughes, 2008). The
185 positive correlation of D ratio and fully vaccination rate became even more evident
186 (R^2= 0.723, $P<10^{-4}$) (Figure 2). Therefore, purifying selection pressure of SARS-CoV-
187 2 spike gene increased as the vaccination coverage rate increased. This suggests that
188 vaccination, in combination with other mitigation strategies, plays an important role
189 in the purifying selection force of SARS-CoV-2 spike protein.
190
191 **Application of neutrality tests and their limitations**
192 Neutrality tests are very powerful tools to study population genetics and viral
193 evolution. However, these tests require careful interpretation due to two known
194 limitations in this study: (1) sampling across countries and the representation of
195 sequences in public databases, (2) the timing of introductions and vaccination
196 campaigns. Previously we have applied various neutrality tests to analyze SARS-
197 CoV-2 transmission and evolution in Diamond Princess cruise (Yeh and Contreras,
198 2021). In contrast to the shipboard quarantine, where the virus spread in a closed
199 environment, many more complicated factors (mitigation efforts, travel, herd
200 immunity, etc.) could influence viral evolution during pandemics among different
201 countries.
202
203 Bhatt et al. have shown that the Tajima D test performs very well with low type I
204 error rate (less than 5%) when the θ value (the substitutions per site of a sequence
205 alignment) is between 0.1 to 10^{-4} in different RNA viruses (Bhatt et al., 2010).

206 Because θ values in this study ranged between 0.0009 to 0.032, the type I error rate
207 is low. The Tajima *D* test is also much less influenced by RNA recombination, which
208 we have reported previously (Yeh and Contreras, 2020, Yeh and Contreras, 2021).
209 Tajima *D* test results depend on the interplay of many parameters: positive and
210 negative selection, population size dynamics, spatial structure and migration,
211 random genetic drift, etc. A key limitation of the Tajima *D* test is the difficulty to
212 separate the effects of each parameter. However, the selection pressure and
213 demographic expansion can be verified in combination with other neutrality tests.
214 Our results of *DH* and Zeng's *E* test suggests that purifying selection, not
215 demographic expansion, was the acting force. Tajima *D* may show negative values
216 due to a recent bottleneck event rather than selection, but the bottleneck effect
217 should impact all types of polymorphism equally (Tajima, 1989). Significant
218 differences between D_{nonsyn} and D_{syn} (either ΔD or D ratio ($D_{\text{nonsyn}}/D_{\text{syn}}$)), which
219 efficiently eliminated the homogenous effects, strongly indicates that purifying
220 selection was at play and is positively corelated with vaccination coverage rate
221 within these samples.
222
223 The inter-species divergence K_a/K_s test has been shown to be poorly equipped to
224 detect positive selection in short time scales for the comparison of young
225 evolutionary lineages like SARS-CoV-2 (Mugal, et al., 2014). Consistent with other
226 analytical attempts, our K_a/K_s data was inconclusive (Table 1). This has been the
227 case with other applications of the K_a/K_s test in SARS-CoV-2 spike gene, which
228 resulted in extreme values and controversial results (Kang et al., 2021),

229

230 In contrast to K_a/K_s, the MK test cannot be used to locate specific sites under
231 selection, although this test enables us to estimate the adaptive substitution rate of a
232 gene (Smith and Eyre-Walker, 2002). Variants of the MK test have been used to
233 analyze the evolution of RNA viruses (Bhatt et al., 2010). Since the MK test relies on
234 the comparison of the ratio of polymorphism to fixed differences of both
235 synonymous and nonsynonymous mutations, selection of an outgroup sequence is
236 critical for this test. The number of fixed synonymous mutations may be
237 underestimated in a rapidly evolving virus like SARS-CoV-2 when a distant outgroup
238 is applied (Baudry and Depaulis, 2003). Since the K_a/K_s or the MK tests may not be
239 the appropriate tests for SARS-CoV-2 delta variants, we only included them to
240 confirm the observation of purifying selection of SARS-CoV-2 spike gene here as
241 well. We also did not include neutrality tests based on haplotype distribution or
242 linkage disequilibrium, because they are expected to be strongly affected by RNA
243 recombination that commonly occurred in SARS-CoV-2 populations (Yeh and
244 Contreras, 2020, Ignatieva et al, 2021, Yeh and Contreras, 2021).

245

246 **Perspectives**

247 More virulent strains of SARS-CoV-2 have emerged with enhanced transmissibility
248 and immune evasion properties. For example, multiple variants have escaped
249 neutralizing antibodies developed to target the spike protein receptor-binding or N-
250 terminal domain (Harvey, et al., 2021). What's more, the case numbers of

251 breakthrough infection caused by the delta variant have increased drastically
252 worldwide (Farinholt et al., 2021).
253
254 Of all countries studied, our observations show that higher vaccination rates
255 corresponded with lower mutation frequencies and higher $D_{\mathrm{nonsyn}}/D_{\mathrm{syn}}$ values for
256 the whole genome and the spike gene. We conclude then, that as the vaccination
257 coverage rate increases, purifying selection forces of nonsynonymous mutations
258 also increase.
259
260 Thus, we recommend that: 1) universal vaccination should be administered as soon
261 as possible to suppress the generation of deadly mutations; 2) mitigations strategies
262 such as personal protection equipment, social distancing, etc., should be continued
263 until full vaccination is reached to prevent viral transmission; 3) genomic
264 surveillance should be undertaken to monitor for new mutations; and 4) more
265 sequence summary statistics of RNA virus evolution are required to facilitate
266 understanding new COVID-19 outbreaks.
267
268
269

Disclosure statement

No potential conflict of interest was reported by the author(s).

Author contributions

All authors contributed to study concept, rationale, and initial manuscript drafts, interpretation of data, and final manuscript preparation. All authors have read and approved the final version of the manuscript.

Funding

No funding

Ethical approval

None declared.

References

Bhatt, B., Katzourakis, A., Pybus, O.G., (2010) Detecting natural selection in RNA virus populations using sequence summary statistics. Infect Genet Evol., *10*, 421-30.

Baudry, E., Depaulis, F. (2003) Effect of misoriented sites on neutrality tests with outgroup. Genetics, *165*, 1619-1622

Chaw, S.M., Tai, J.H., Chen, S.L., Hsieh, C.H., Chang, S.Y., Yeh, S.H., Yang, W.S., Chen, P.J., Wang, H.Y. (2020) The origin and underlying driving forces of the SARS-CoV-2 outbreak. J Biomed Sci., *27*, 73.

Farinholt, T., Doddapaneni, H., Qin, X., Menon, V., Meng, Q., Metcalf, G., Chao, H., Gingras M., Farinholt, P., Agrawal, C., Muzny, D.M., Piedra, P.A., Gibbs, R.A., Petrosino, J. (2021) Transmission event of SARS-CoV-2 Delta variant reveals

296 multiple vaccine breakthrough infections. medRxiv doi:
297 10.1101/2021.06.28.21258780.
298 Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M.,
299 Ludden, C., Reeve, R., Rambaut, A., COVID-19 Genomics UK (COG-UK) Consortium,
300 Peacock S.H., Robertson D.L. (2021) SARS-CoV-2 variants, spike mutations and
301 immune escape. Nat Rev Microbiol. ,*19*, 409-424.
302 Hahn, M.W., Rausher, M.D., Cunningham, C.W. (2002) Distinguishing between
303 selection and population expansion in an experimental lineage of bacteriophage T7.
304 Genetics, *161*, 11-20.
305 Hughes, A.L., (2005) Evidence for Abundant Slightly Deleterious Polymorphisms in
306 Bacterial Populations. Genetics, 169, 533-538.
307 Hughes, A.L., Friedman, R., Rivailler, P., French, J.O., (2008) Synonymous and
308 nonsynonymous polymorphisms versus divergences in bacterial genomes. Mol Biol
309 Evol. ,*25*, 2199–2209.
310 Ignatieva, A., Hein, J., Jenkins, P.A, (2021) Evidence of ongoing recombination in
311 SARS-CoV-2 through genealogical reconstruction. bioRxiv,
312 https://doi.org/10.1101/2021.01.21.427579
313 Kang, L., He, G., Sharp, A.K., Wang, X., Brown, A.M., Michalak. P., Weger-Lucarelli, J.
314 (2021) A selective sweep in the Spike gene has driven SARS-CoV-2 human
315 adaptation. Cell, *184*, 4392-4400.
316 Kirzinger, A., Sparks, G., Brodie, M. (2021) KFF COVID-19 vaccine monitor-Rural
317 America. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-
318 vaccine-monitor-in-their-own-words-six-months-later/ Accessed 13 July 2021.

319 Mugal, C.F., Wolf, J.B., Kaj, I. (2013) Why time matters: codon evolution and the
320 temporal dynamics of dN/dS. Mol Biol Evol., *31*,212–231.
321 Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-
322 Onsins, S.E., Sánchez-Gracia, A. (2017) DnaSP 6: DNA Sequence Polymorphism
323 Analysis of Large Data Sets. Mol Biol Evol., *34*, 3299–3302.
324 Roy, C., Mandal, S.M., Mondal, S.K., Mukherjee, S., Mapder, T., Ghosh, W., Chakraborty,
325 R. (2020) Trends of mutation accumulation across global SARS-CoV-2 genomes:
326 Implications for the evolution of the novel coronavirus. Genomics ,*112*, 5331-5342.
327 Schmelz, K., Bowles, S., (2021) Overcoming COVID-19 vaccination resistance when
328 alternative policies affect the dynamics of conformism, social norms, and crowding
329 out. Proc Natl Acad Sci U S A ,*118*, e2104912118.
330 Smith, N.G., Eyre-Walker, A., (2002) Adaptive protein evolution in *Drosophila*.
331 Nature, *415*, 1022–1024.
332 Tajima, F. (1989) Statistical method for testing the neutral mutation hypothesis by
333 DNA polymorphism. Genetics, *123*, 585–595.
334 Yeh, T.Y., Contreras, G.P. (2020) Emerging viral mutants in Australia suggest RNA
335 recombination event in the SARS-CoV-2 genome. Med J Aust ,*213*, 44-44.e1.
336 Yeh, T.Y., Contreras, G.P. (2021) Viral transmission and evolution dynamics of SARS-
337 CoV-2 in shipboard quarantine. Bull World Health Organ., *99*, 486–495.
338 Zeng, K., Fu, Y.X., Shi, S., Wu, C.I. (2006) Statistical tests for detecting positive
339 selection by utilizing high-frequency variants. Genetics, *174*, 1431-1439.
340

341 **Figure Legend**

342 Figure 1. Correlation between full vaccinated rate

343 (https://ourworldindata.org/covid-vaccinations) and mutation frequency (*Mf*) of

344 the full genome (A) or the spike gene (B) of SARS-CoV-2 delta variants from June 20

345 to July 3 2021 in 20 countries: Australia (AUS), France (FRA), Germany (GER),

346 Indonesia (IDA), India (IND), Ireland (IRL), Israel (ISR), Italy (ITA), Japan (JPN),

347 Mexico (MEX), Netherland (NED), Norway (NOR), Portugal (POR), Singapore (SGP),

348 Spain (ESP), Switzerland (SUI), Sweden (SWE), Turkey (TUR), United States (USA),

349 and UK. Logarithmic regression (solid) line was draw based on 16 countries (pink)

350 with a calculated 95% confidence interval (dashed lines). Japan, Switzerland, USA,

351 and Australia are labeled in different colors as outliers.

352 Figure 2. Correlation between full vaccinated rate and *D* ratio ($D_{\mathrm{nonsyn}}/D_{\mathrm{syn}}$) of the

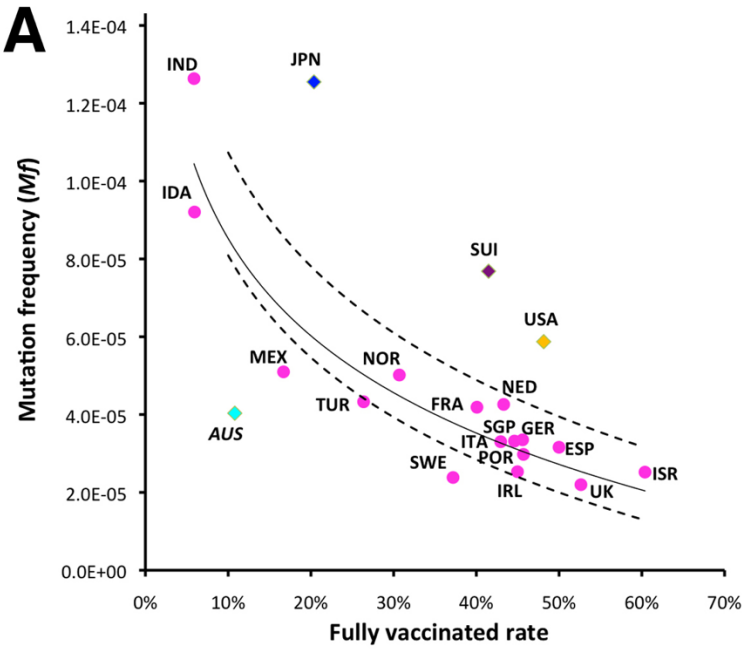
353 spike gene of SARS-CoV-2 delta variants. Exponential regression (solid) line was

354 draw based on 19 countries (pink square) with a calculated 95% confidence interval

355 (dashed lines). Australia (AUS) is labeled as outliers.

356

Figure 1.


A
Mutation frequency (Mf)
1.4E-04
1.2E-04
1.0E-04
8.0E-05
6.0E-05
4.0E-05
2.0E-05
0.0E+00
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate
IND
JPN
IDA
SUI
USA
MEX
NOR
NED
TUR
FRA
SGP
GER
AUS
ITA
ESP
POR
SWE
ISR
IRL
UK

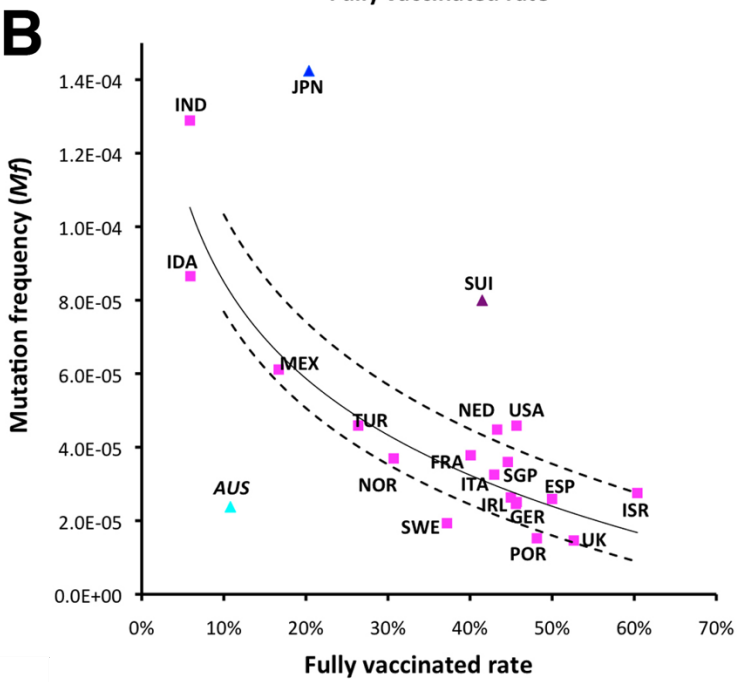
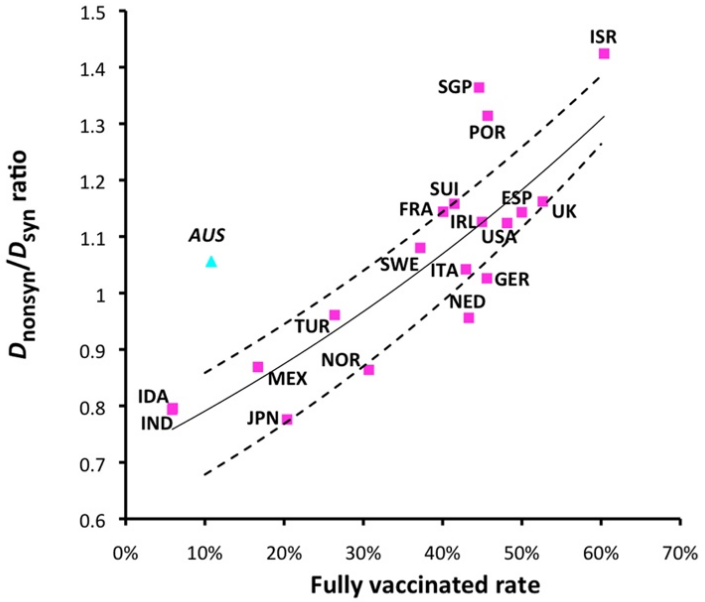
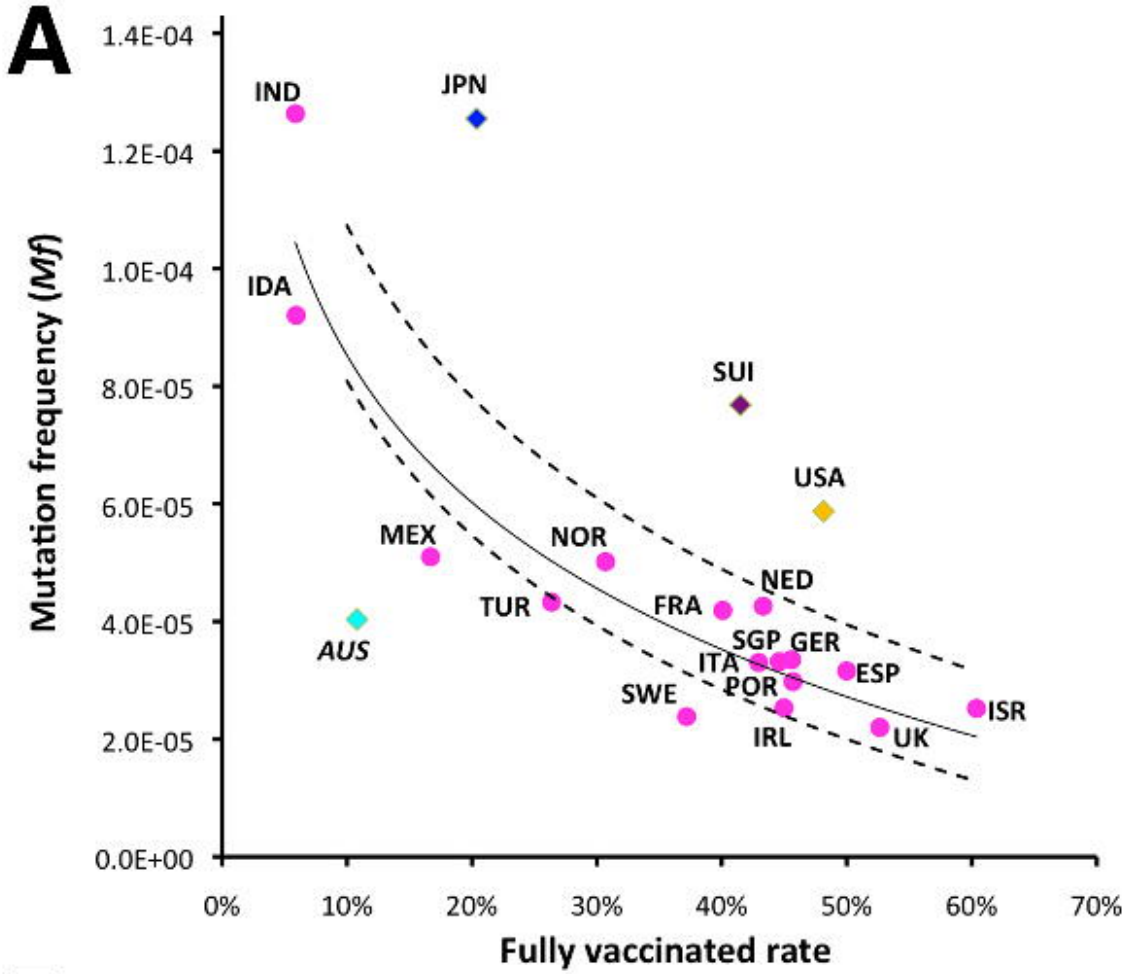

B
Mutation frequency (Mf)
1.4E-04
1.2E-04
1.0E-04
8.0E-05
6.0E-05
4.0E-05
2.0E-05
0.0E+00
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate
JPN
IND
IDA
SUI
MEX
NED
USA
TUR
FRA
SGP
ITA
ESP
AUS
NOR
IRL
ISR
GER
SWE
UK
POR

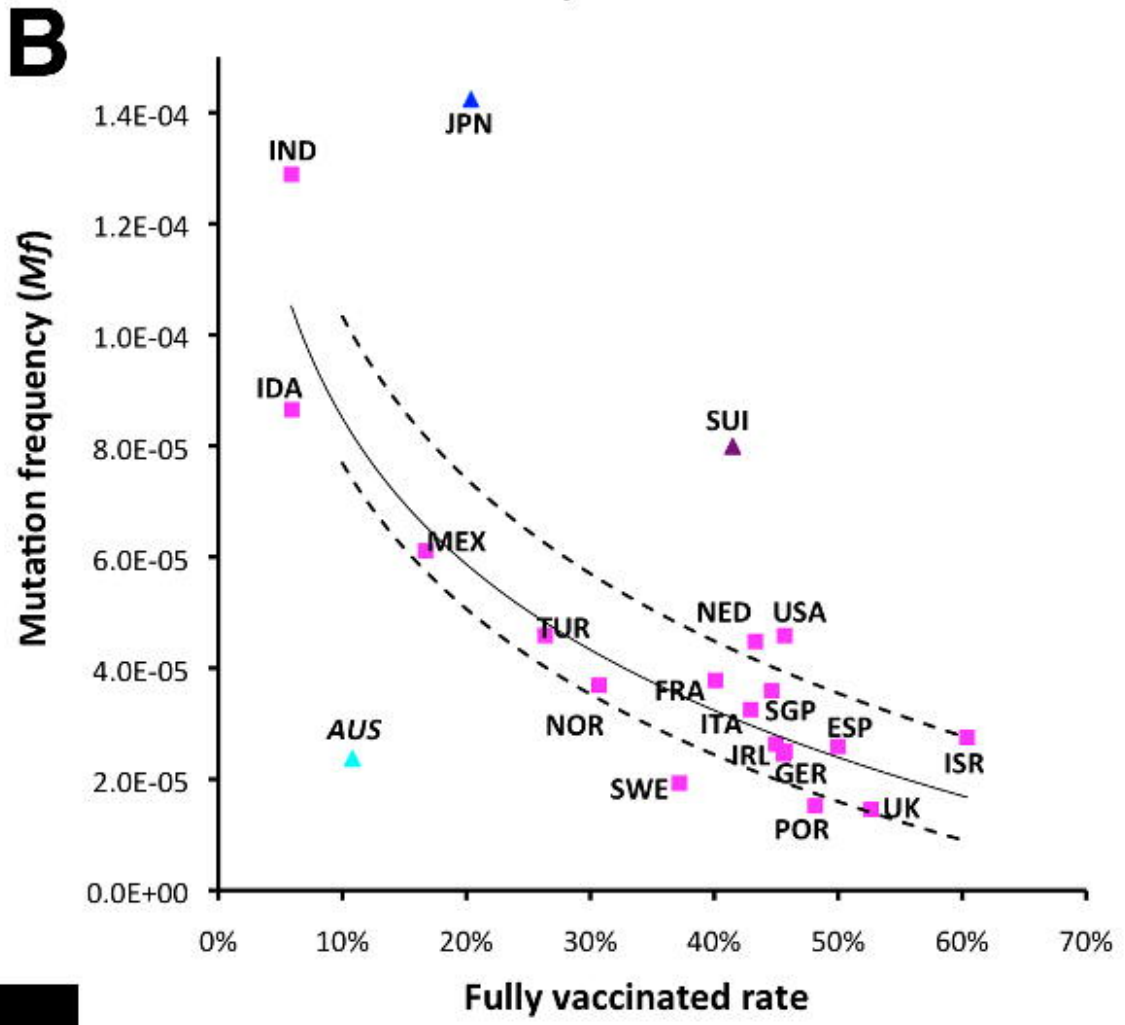
Figure 2.


1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate
D_{nonsyn}/D_{syn} ratio
ISR
SGP
POR
SUI
FRA
ESP
UK
IRL
USA
AUS
SWE
ITA
GER
NED
TUR
NOR
MEX
IDA
IND
JPN

365 Table 1. Neutrality analysis of SARS-CoV-2 delta variants of the full genome and the viral spike gene
366 from June 20 to July 3, 2021 in 20 countries. The statistical significance was estimated using 10,000
367 coalescent simulations in DNASP6. Not significant values (P>0.05) are indicated in italic.
368

Country	Tajima's *D*	*DH*	Zeng's *E*	*DH*	Zeng's *E*	Tajima's *D*	ΔD (D_{Nonsyn} -D_{syn})	*D* ratio (D_{Nonsyn} /D_{syn})	Ka /Ks	Neutrality index
Israel	-2.401	-2.539	*-0.040*	-4.468	*1.792*	-2.065	0.611	1.424	0.055	13.209
France	-2.518	-2.482	*-0.278*	-2.973	*0.176*	-2.482	0.307	1.144	0.055	15.612
Germany	-2.800	-2.703	*-0.370*	-3.831	*0.655*	-2.661	0.065	1.026	0.034	10.938
Portugal	-2.645	-2.269	*-0.558*	-3.16	*0.402*	-2.395	0.593	1.314	0.058	8.393
Spain	-2.660	-2.61	*-0.283*	*-1.606*	*-0.943*	-2.530	0.306	1.143	0.065	12.986
Australia	-2.578	-1.914	*-0.626*	*-1.949*	*-0.078*	-2.064	0.091	1.056	0.052	10.609
USA	-2.657	-2.207	*-0.594*	-4.551	*1.306*	-2.580	0.290	1.124	0.015	14.467
Ireland	-2.833	-2.069	*-0.868*	-2.337	*-0.464*	-2.659	0.289	1.126	0.055	12.385
Italy	-2.533	-2.341	*-0.388*	-3.691	*0.73*	-2.51	0.095	1.042	0.045	17.699
Turkey	-2.451	-2.638	*-0.252*	-3.748	*0.916*	-2.398	-0.09	0.961	0.044	16.292
Indonesia	-2.339	-2.116	*-0.211*	*-1.35*	*-0.342*	*-1.773*	-0.381	0.796	0.055	11.236
Japan	-2.523	-1.801	*-0.675*	-3.697	*1.095*	-2.459	-0.582	0.776	0.052	5.068
Mexico	-2.522	-2.532	*-0.166*	-3.111	*0.444*	-2.405	-0.307	0.869	0.073	11.340
Norway	*-1.566*	*-1.177*	*-0.370*	*-0.682*	*-0.732*	*-1.562*	0.359	1.307	0.054	9.059
Swistzerland	-2.560	-2.319	*-0.279*	-4.591	*1.756*	-2.494	0.333	1.158	0.045	10.81
Sweden	-2.473	-2.295	*-0.383*	-4.275	*1.101*	-2.565	0.182	1.08	0.021	11.19
Netherlands	-2.372	-2.327	*-0.231*	-3.811	*0.893*	-2.48	-0.107	0.956	0.049	10.64
Singapore	-2.199	-2.791	*0.482*	-2.538	*0.507*	-1.909	0.500	1.364	0.053	9.059
India	-2.878	-2.470	*-0.469*	-3.961	*0.884*	-2.826	-0.073	0.793	0.051	7.102
UK	-2.748	-2.546	*-0.540*	-3.299	*0.168*	-2.663	0.355	1.162	0.013	32.537


A
Mutation frequency (Mf)
1.4E-04
1.2E-04
1.0E-04
8.0E-05
6.0E-05
4.0E-05
2.0E-05
0.0E+00
IND
JPN
IDA
SUI
USA
MEX
NOR
NED
TUR
FRA
AUS
SGP
GER
ITA
ESP
POR
SWE
ISR
IRL
UK
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate


B
Mutation frequency (Mf)
1.4E-04
1.2E-04
1.0E-04
8.0E-05
6.0E-05
4.0E-05
2.0E-05
0.0E+00
JPN
IND
IDA
SUI
MEX
NED
USA
TUR
FRA
SGP
AUS
NOR
ITA
ESP
JRL
ISR
GER
SWE
UK
POR
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate

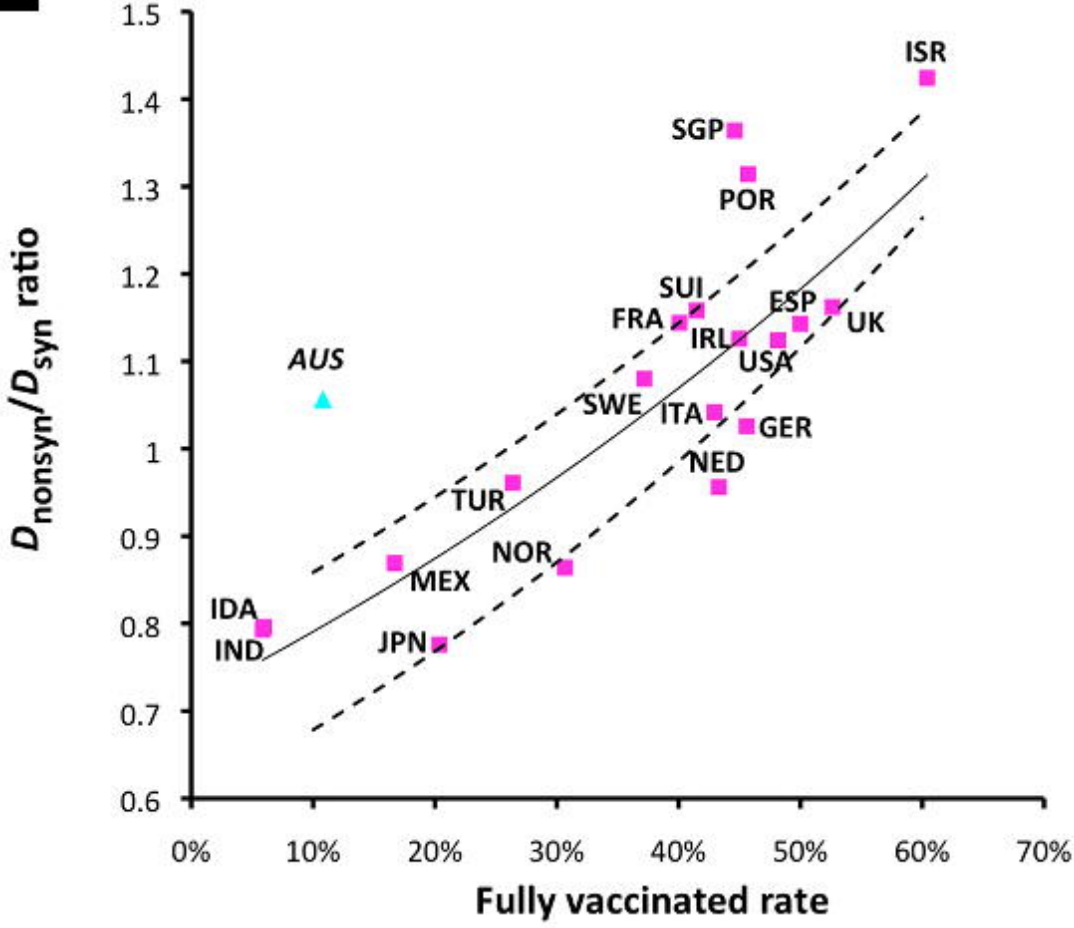

1.5
1.4
1.3
1.2
1.1
1
0.9
0.8
0.7
0.6
D_{nonsyn}/D_{syn} ratio
0%
10%
20%
30%
40%
50%
60%
70%
Fully vaccinated rate
ISR
SGP
POR
SUI
FRA
ESP
UK
IRL
USA
AUS
SWE
ITA
GER
NED
TUR
NOR
MEX
IDA
IND
JPN

Table 1. Neutrality analysis of SARS-CoV-2 delta variants of the full genome and the viral spike gene from June 20 to July 3, 2021 in 20 countries. The statistic significance was estimated using 10,000 coalescent simulations in DNASP6. Not significant values (P>0.05) are indicated in italic.

	Full genome			Spike gene						
Country	**Tajima's *D***	***DH***	**Zeng's *E***	***DH***	**Zeng's *E***	**Tajima's *D***	**ΔD (D_{Nonsyn} -D_{syn})**	***D* ratio (D_{Nonsyn} /D_{syn})**	**Ka /Ks**	**Neutrality index**
Israel	-2.401	-2.539	*-0.040*	-4.468	*1.792*	-2.065	0.611	1.424	0.055	13.209
France	-2.518	-2.482	*-0.278*	-2.973	*0.176*	-2.482	0.307	1.144	0.055	15.612
Germany	-2.800	-2.703	*-0.370*	-3.831	*0.655*	-2.661	0.065	1.026	0.034	10.938
Portugal	-2.645	-2.269	*-0.558*	-3.16	*0.402*	-2.395	0.593	1.314	0.058	8.393
Spain	-2.660	-2.61	*-0.283*	*-1.606*	*-0.943*	-2.530	0.306	1.143	0.065	12.986
Australia	-2.578	-1.914	*-0.626*	*-1.949*	*-0.078*	-2.064	0.091	1.056	0.052	10.609
USA	-2.657	-2.207	*-0.594*	-4.551	*1.306*	-2.580	0.290	1.124	0.015	14.467
Ireland	-2.833	-2.069	*-0.868*	-2.337	*-0.464*	-2.659	0.289	1.126	0.055	12.385
Italy	-2.533	-2.341	*-0.388*	-3.691	*0.73*	-2.51	0.095	1.042	0.045	17.699
Turkey	-2.451	-2.638	*-0.252*	-3.748	*0.916*	-2.398	-0.09	0.961	0.044	16.292
Indonesia	-2.339	-2.116	*-0.211*	*-1.35*	*-0.342*	*-1.773*	-0.381	0.796	0.055	11.236
Japan	-2.523	-1.801	*-0.675*	-3.697	*1.095*	-2.459	-0.582	0.776	0.052	5.068
Mexico	-2.522	-2.532	*-0.166*	-3.111	*0.444*	-2.405	-0.307	0.869	0.073	11.340
Norway	*-1.566*	*-1.177*	*-0.370*	*-0.682*	*-0.732*	*-1.562*	0.359	1.307	0.054	9.059
Swistzerland	-2.560	-2.319	*-0.279*	-4.591	*1.756*	-2.494	0.333	1.158	0.045	10.81
Sweden	-2.473	-2.295	*-0.383*	-4.275	*1.101*	-2.565	0.182	1.08	0.021	11.19
Netherlands	-2.372	-2.327	*-0.231*	-3.811	*0.893*	-2.48	-0.107	0.956	0.049	10.64
Singapore	-2.199	-2.791	*0.482*	-2.538	*0.507*	-1.909	0.500	1.364	0.053	9.059
India	-2.878	-2.470	*-0.469*	-3.961	*0.884*	-2.826	-0.073	0.793	0.051	7.102
UK	-2.748	-2.546	*-0.540*	-3.299	*0.168*	-2.663	0.355	1.162	0.013	32.537