medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

SARS-CoV-2 Variant Under Investigation 202012/01 has more than twofold replicative advantage

Frederic Grabowski[1], Grzegorz Preibisch[2], Marek Kochańczyk[1], and Tomasz Lipniacki[1,*]

[1]Institute of Fundamental Technological Research,
Polish Academy of Sciences, 02-106 Warsaw, Poland

[2]Inter-Faculty Individual Studies in Mathematics and Natural Sciences,
The MISMaP College, University of Warsaw, 02-087 Warsaw, Poland

*Corresponding author's e-mail: `tlipnia@ippt.pan.pl`

Abstract

The novel SARS-CoV-2 Variant Under Investigation VUI-202012/01 (also known as B.1.1.7), first collected on September 20, 2020, in Kent, is a rapidly growing lineage that in the first half of December 2020 constitutes more than 1/3 of all SARS-CoV-2 genomes sequenced in England. Based on GISAID data we have shown nearly exponential growth of the VUI-202012/01 to non-VUI-202012/01 genomes ratio in the five-week period since October 19 till November 22, 2020, with weekly growth rate of 2.31 [95% CI: 2.08–2.57]. Assuming the serial interval of 6.73 days we estimated the replicative advantage of VUI-202012/01 lineage as $2.31^{6.73/7} = 2.24$ [95% CI: 2.03–2.48]. Such significant replicative advantage and the fact that London serves as major international transportation hub suggest that the VUI-202012/01 strain will likely become globally dominant, hindering containment of the COVID-19 epidemics prior to massive vaccinations.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

The earliest two genomes belonging to the novel SARS-CoV-2 Variant Under Investigation (VUI)-202012/01, also known as B.1.1.7 lineage, were collected on September 20 and 21, 2020, in Kent and London (GISAID sequence accession numbers: `EPI_ISL_601443` and `EPI_ISL_581117`, respectively). The lineage started to grow rapidly in mid-October 2020, constituting in the first half of December 2020 more than 1/3 of all SARS-CoV-2 genomes sequenced in England. The lineage is defined by nine spike protein mutations (deletions 69-70HV, 145V, substitutions N501Y, A570D, D614G, P681H, T716I, S982A, D1118H).[1,2] Deletion 69-70HV that is characteristic (but not unique) to VUI-202012/01 causes that the spike-gene is not detected by the dPCR probe used in the English diagnostic system (Spike Gene Target Failure, SGTF). As VUI-202012/01 strain has a high multiplicative potential, it has become the prevalent del-69-70HV strain. SGTF can be used to provide an estimate of the prevalence of the VUI-202012/01 genome.[3] Based on dPCR tests, Public Health England group determined multiplicative advantage of VUI-202012/01 in NHS England STP areas in subsequent weeks 44–48 of 2020. As an average for the analyzed areas the authors obtained the ratio of reproduction numbers equal 1.47 (95% CI: 1.34–1.59). Also, as an average over three regions, Davies *et al.*[4] estimated that VUI-202012/01 strain has 56% higher transmissibility (95% CI: 50%–74%).

Here, we propose a more direct approach in which we use England-aggregated data from GISAID and calculate the ratio of VUI-202012/01 to non-VUI-202012/01 genomes in subsequent weeks. We observe that this ratio grows nearly exponentially in weeks 43 to 47, for which the data are close to complete and for which the number of collected VUI-202012/01 genomes exceeds 27. In this period, about 5% of cases in England had sequenced genomes. Based on the weekly growth rate of the VUI-202012/01 to non-VUI-202012/01 genomes we estimated the ratio of the replication numbers of VUI-202012/01 to non-VUI-202012/01 strains.

---

[1] GISAID. *UK reports new variant, termed VUI 202012/01*. https://www.gisaid.org/references/gisaid-in-the-news/uk-reports-new-variant-termed-vui-20201201. Accessed: 2020-12-27.

[2] CoG-UK. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations*. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. Accessed: 2020-12-27.

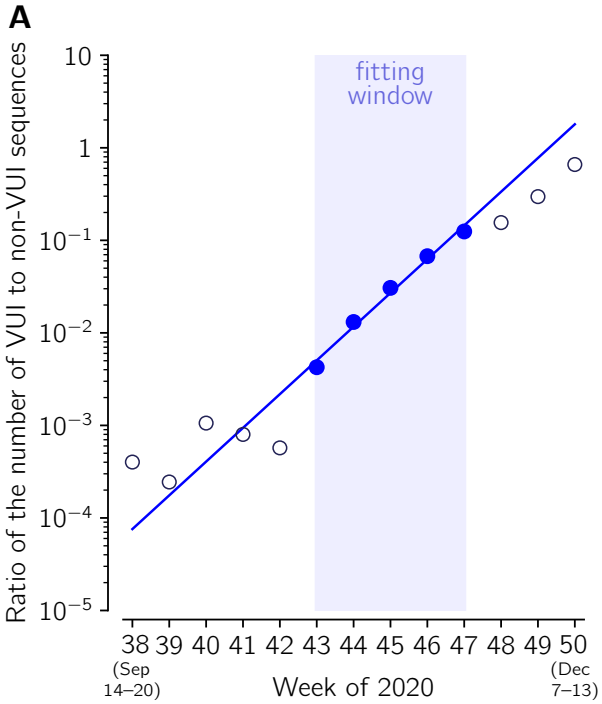
[3] England, P. H. *Investigation of novel SARS-COV-2 variant, Variant of Concern 202012/01*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf. Accessed: 2020-12-21.

[4] Davies, N. G. *et al.* Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *medRxiv*. https://dx.doi.org/10.1101/2020.12.24.20248822 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

To estimate the replicative advantage of VUI-202012/01 strain over non-VUI-202012/01 strains we use England-aggregated SARS-CoV-2 genomic data submitted to GISAID till 8pm UTC, December 26, 2020. We calculate the growth rate of the VUI-202012/01 to non-VUI-202012/01 fraction f in the five-week period since October 19 till November 22, 2020, in which the growth is nearly exponential. By using week intervals we remove effects of weekly patterns associated with irregular genome collection. In weeks 38 till 42 there were less than 5 VUI-202012/01 genomes collected per week. In week 43 the number of the collected VUI-202012/01 genomes rose to 28, beginning a regular exponential growth till week 47, Fig. 1. In week 48 this trend appears to be terminated, possibly due to the incompleteness of data. In weeks 43 till 47, fraction f exhibits weekly growth rate of 2.31 [95% CI: 2.08–2.57].



B

Week	Genomes		Positive tests	
	VUI	non-VUI	count	sequenced
38	1	2490	21 780	11.4%
39	1	4088	33 818	12.4%
40	3	2832	57 124	5.0%
41	4	5006	83 596	6.0%
42	3	5247	98 407	5.3%
43	28	6583	127 416	5.2%
44	68	5186	138 784	3.8%
45	252	8225	136 626	6.2%
46	532	7892	159 307	5.3%
47	509	4081	126 580	3.6%
48	337	2167	90 647	2.8%
49	374	1260	88 174	1.9%
50	1463	2217	105 195	3.5%

Figure 1: The ratio of VUI-202012/01 to non-VUI-202012/01 genomes collected in weeks 38–50, 2020, based on GISAID data submitted till Dec 26, 2020. The trend line in (**A**) is fitted to data points from weeks 43–47 (blue), in which the number of collected VUI-202012/01 genomes exceeds 27 and the total number of genomes exceeds 4500 per week. The weekly growth rate is 2.31 [95% CI: 2.08–2.57]. The number of positive cases in (**B**) is given after https://coronavirus.data.gov.uk/details/cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To estimate the ratio of reproduction numbers $\mathcal{R}_t^{\text{VUI}}/\mathcal{R}_t^{\text{non-VUIs}}$ of VUI-202012/01 and non-VUI-202012/01 strains we assumed that both strains have the same serial interval of 6.73 days.[5] Then $\mathcal{R}_t^{\text{VUI}}/\mathcal{R}_t^{\text{non-VUIs}} = 2.31^{6.73/7} = 2.24$ [95% CI: 2.03–2.48].

Discussion

We have estimated the replicative advantage of VUI-202012/01 strain, that is, the ratio of reproduction numbers of VUI-202012/01 and non-VUI-202012/01 strains of SARS-CoV-2, as $\mathcal{R}_t^{\text{VUI}}/\mathcal{R}_t^{\text{non-VUIs}} = 2.24$ [95% CI: 2.03–2.48]. Importantly, since our estimate is based the growth rate of the ratio of VUI-202012/01 to non-VUI-202012/01 genomes in subsequent weeks, it is not influenced by protective measures that modulate the reproduction numbers of VUI-202012/01 and non-VUI-202012/01 strains in the considered period. Our estimation is based on the yet incomplete data and several assumptions discussed below.

1. We used data submitted to GISAID till December 8pm UTC, 26, 2020 aggregated over whole England. In order to check to what extent our estimation is influenced by incompleteness of the data we repeat our estimation using only data submitted to GISAID till December 16, 2020, that contains about 200 VUI-202012/01 genomes (collected in weeks 43–48) less than more complete dataset available on December 26, 2020 (1179 *vs.* 1389). In this case we obtain $\mathcal{R}_t^{\text{VUI (Dec16)}}/\mathcal{R}_t^{\text{non-VUIs (Dec16)}} = 2.29$ (95% CI: 2.11–2.48), close to the value obtained for December 26 equal 2.24.

2. We used data aggregated over England, not the whole UK, as in weeks 43–48 there were only 19 non-England UK cases of the VUI. Nevertheless, when the whole UK is considered, one obtains $\mathcal{R}_t^{\text{VUI (UK)}}/\mathcal{R}_t^{\text{non-VUIs (UK)}} = 2.30$ (95% CI: 2.04–2.59), close to value obtained for England, equal 2.24.

3. We think that the England-aggregated data-based approach gives a more accurate estimation than the approach in which replicative advantage is calculated independently for NHS England STP areas and then averaged. This is because, as implicated by Fig. 2, the

[5] Kochańczyk, M. *et al.* Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}_0$ higher than initially estimated. *Royal Society Open Science* **7**, 200786. https://dx.doi.org/10.1098/rsos.200786 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

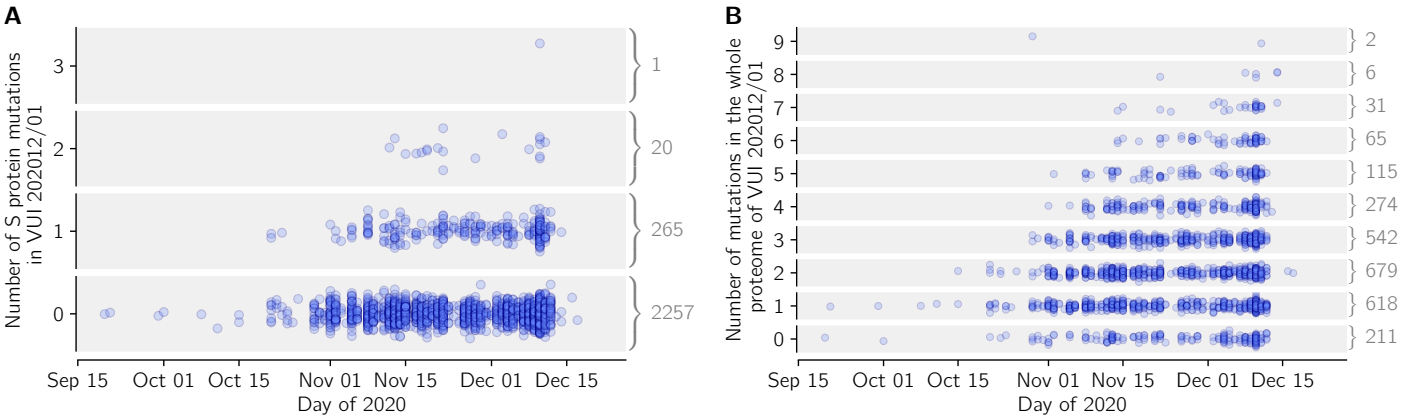


Figure 2: Mutations in VUI-202012/01 lineage with respect to the first VUI-202012/01 genome collected in September 20, 2020 (GISAID sequence accession EPI_ISL_601443): (A) mutations in the spike protein, (B) mutations in the whole proteome. On the right-hand side of each panel the number of the collected genomes with a given Levenshtein editing distance from the EPI_ISL_601443 genome is given (gray numbers next to a brace). Only properly annotated genomes without outliers are included.

VUI-202012/01 strain evolves and has already a complex phylogenetic tree.[6] It is therefore likely that some sub-strains have a lower or even no replicative advantage and will be extinct in the course of evolution, and finally the sub-strain with the highest replicative advantage will dominate. By using the aggregated approach we estimate the replicative advantage of dominating sub-strain(s).

4. We have estimated the weekly growth of fraction f (ratio of the VUI-202012/01 and non-VUI-202012/01 strains abundances) using only 5 data points corresponding to weeks 43–48. Since the count of VUI-202012/01 genomes is relatively small, it could happen that the nearly linear dependence was obtained by chance and thus the classically calculated confidence interval is too narrow. To check for this we estimated the confidence interval assuming that VUI-202012/01 genomes are drawn randomly from a binomial distribution (see Methods). We found that such theoretically estimated CI, 2.06–2.35, is somewhat narrower than the range 2.03–2.48 obtained calculating the standard error of the slope. This shows that our result $\mathcal{R}_t^{\text{VUI}}/\mathcal{R}_t^{\text{non-VUIs}}$ = 2.24 [95% CI: 2.03–2.48] with relatively narrow confidence interval was not obtained by chance.

[6] CoG-UK. *Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563. Accessed: 2020-12-27.

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

5. To convert the weekly growth of f to the ratio of respective reproduction numbers we assumed that both strains have the same serial interval equal 6.73 days. This calculation is approximate as the serial interval is not a number but follows hypoexponential distribution[7]. Additionally, although no current data indicate this, it may happen that the faster spread of VUI-202012/01 strain partially results from the shorter generation time.

In conclusion, we have shown that new VUI-202012/01 strain has more than twofold replicative advantage over non-VUI-202012/01 strains of SARS-CoV-2. Since London is a major transportation hub it is plausible that this strain has already spread over Europe and USA, and will became the globally dominant strain. As of December 27, 2020, twelve VUI-202012/01 strain genomes were reported from Italy, 9 from Denmark, 7 from Ireland, 6 from Japan, 4 from Israel, 3 from Australia, 1 from Singapore, and 1 Hong Kong. Spread of the faster-replicating VUI-202012/01 strain may hinder the efforts to contain the COVID-19 epidemics prior to massive vaccinations.

Methods

The 95% confidence interval for growth rate of f (the VUI-202012/01 to non-VUI-202012/01 abundances fraction) is calculated from the standard error of the slope, as 1.96 $\times$ (standard error). We have also estimated CI by assuming that the number of sequenced VUI-202012/01 genomes is drawn from the binomial distribution with success probability $p_i = n_{\text{VUI}}/(n_{\text{VUI}} + n_{\text{non-VUI}})$, where n_{VUI} and $n_{\text{VUI}} + n_{\text{non-VUI}}$ are the numbers of VUI-202012/01 and all genomes collected in the ith week. For weeks 43 till 47, from such defined binomial distribution we draw 10^5 numbers of VUI-202012/01 genomes, and for each set of numbers we fit the intercept and the weekly growth rate of VUI-202012/01 to non-VUI-202012/01 fraction f. In the analysis of VUI-202012/01 strain genome evolution (Fig. 2) only confidently determined genome sequences (without letters other than ATCG) and without outliers having more than 15 mutations (36 such outliers were excluded) are considered.

[7] Kochańczyk, M. *et al.* Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}_0$ higher than initially estimated. *Royal Society Open Science* **7,** 200786. https://dx.doi.org/10.1098/rsos.200786 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.12.28.20248906; this version posted January 4, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Acknowledgments

This study was supported by the National Science Centre (Poland) grant 2018/29/B/NZ2/00668.

Author contributions

F.G. conceived study, performed data analysis and prepared figures; G.P. performed analysis and prepared figures; M.K prepared figures and wrote the manuscript; T.L. conceived study and wrote the manuscript. All authors gave final approval for publication.

Corresponding authors

Correspondence to Tomasz Lipniacki.

Competing interests

The authors have no competing interests.

Data availability

All data used in this study are referenced.