

COVID-19 UK introductions

Importation summary

Louis du Plessis

Last modified: 07 Dec 2020

Contents

1	Summary	2
1.1	Input	2
1.2	Output	2
2	Genomes compared to cases and estimated infections	3
3	Arrivals and estimated infections	7
4	EII curves	10
5	Lineage importation distribution (shifted TMRCA distribution)	12
6	Import breakdown	17
7	Session info	22

1 Summary

This notebook plots figures about the dataset, infections in different countries and inbound travellers. It also applies the importation lag model to the UK transmission lineage TMRCAs and plots figures with lineage importations.

1.1 Input

- Metadata table (in `inputpath`).
- Cluster statistics for MCC trees (in `outputpath`):
- `clusters_DTA_MCC_0.5.csv`
- `clusterSamples_DTA_MCC_0.5.csv`
- Files with estimated numbers of infections, inbound travellers and calculated EII curves (in `epipath`):
 - `estimated-daily-infections.csv`
 - `estimated-arrivals.csv`
 - `estimated-introduction-index-31.csv` (assuming 31% asymptomatic)
- Positive cases in the UK (in `casefile`)
- Estimated infections across Europe, from Flaxman et al. (in `infectionfile`)

1.2 Output

- Figures and tables.
- Cluster statistics with estimated lineage importation times (as a `.csv` file).

2 Genomes compared to cases and estimated infections

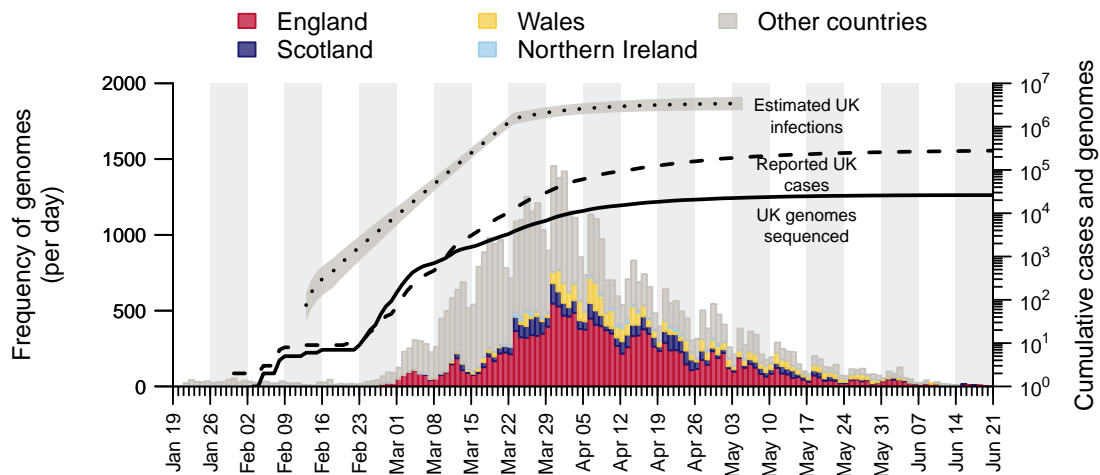


Figure 1: Collection dates of the 50,887 genomes analysed here (left-hand axis). Genomes are coloured by sampling location (England=red, Scotland=dark blue, Wales=yellow, Northern Ireland=light blue, elsewhere=grey). The solid line shows the cumulative number of UK virus genomes (right-hand axis). The dashed and dotted lines show, respectively, the cumulative number of laboratory-confirmed UK cases (by specimen date) and the estimated number of UK infections (Flaxman et al. 2020; grey shading=95% CI; right-hand axis). Due to retrospective screening, the cumulative number of genomes early in the epidemic exceeds that of confirmed cases.

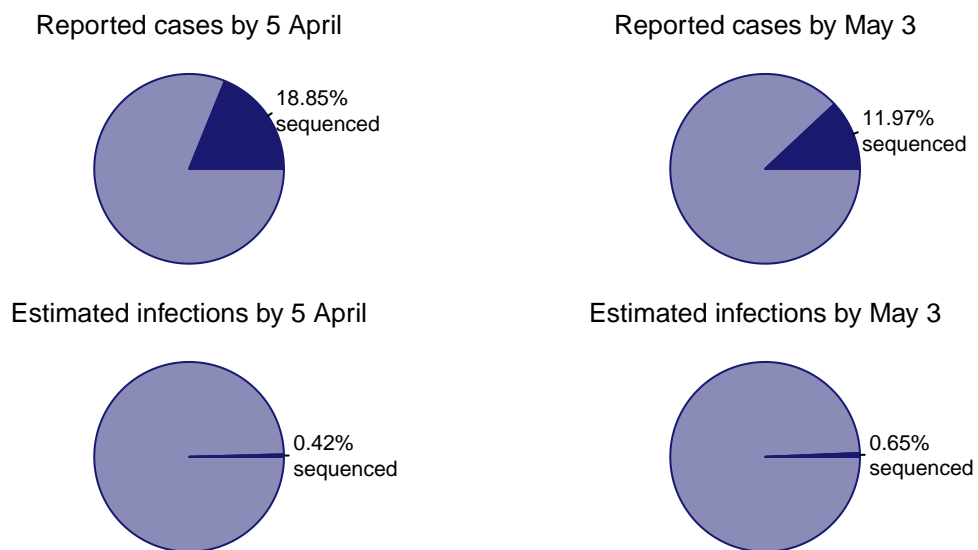


Figure 2: Top row: Proportions of laboratory-confirmed UK cases sequenced by 5 April and 3 May, respectively. Bottom row: Proportions of the estimated number of UK infections (estimate from Flaxman et al. 2020) sequenced by 5 April and 3 May, respectively.

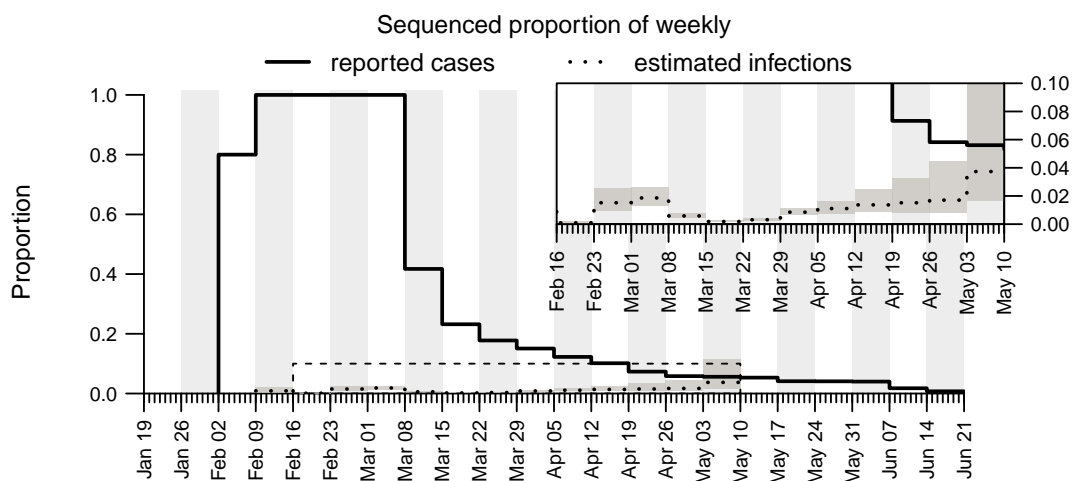


Figure 3: Proportion of weekly reported UK cases (solid line) and estimated infections (dotted line; Flaxman et al. 2020) included in our genome sequence dataset.

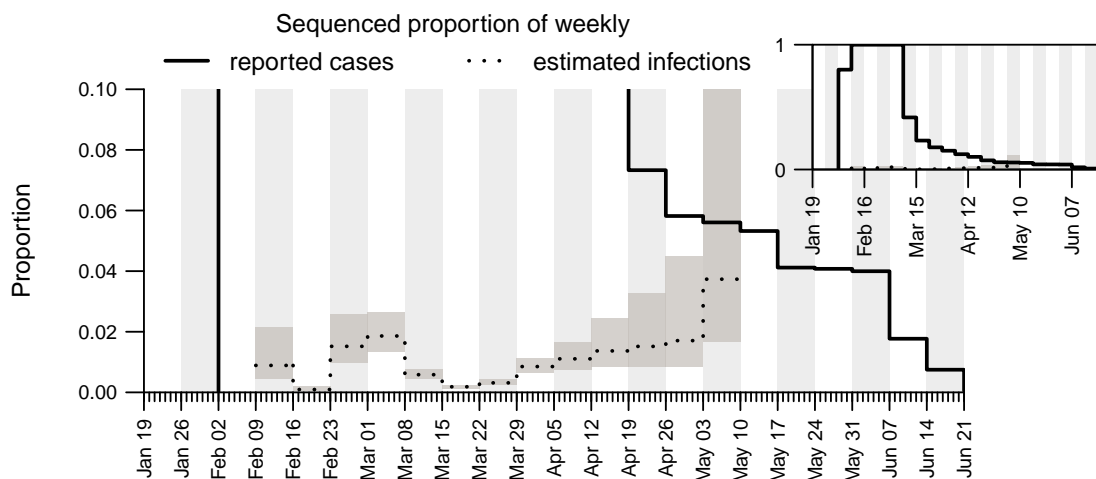


Figure 4: Proportion of weekly reported UK cases (solid line) and estimated infections (dotted line; Flaxman et al. 2020) included in our genome sequence dataset.

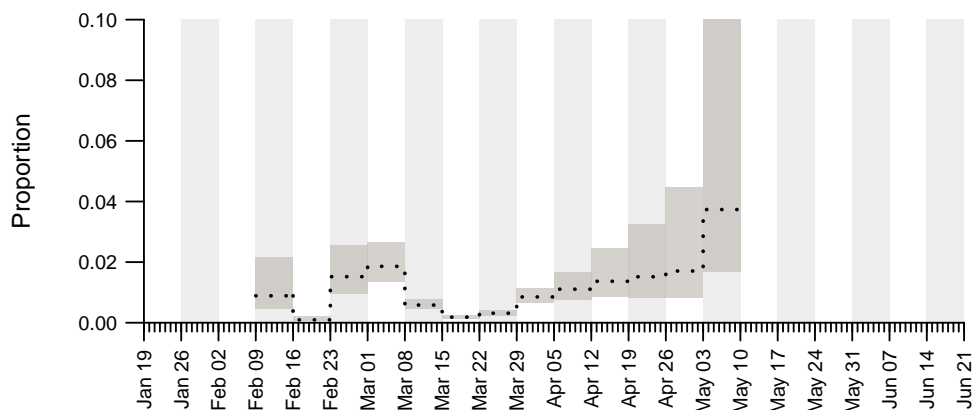


Figure 5: Proportion of weekly estimated UK infections (Flaxman et al. 2020) included in our genome sequence dataset.

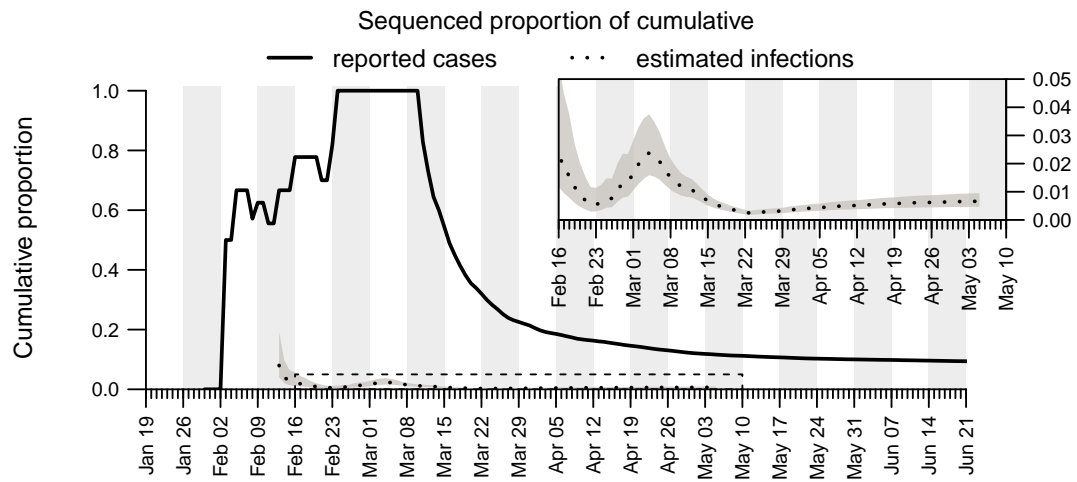


Figure 6: Proportion of the cumulative weekly reported UK cases (solid line) and estimated infections (dotted line; Flaxman et al. 2020) included in our genome sequence dataset over time.

3 Arrivals and estimated infections

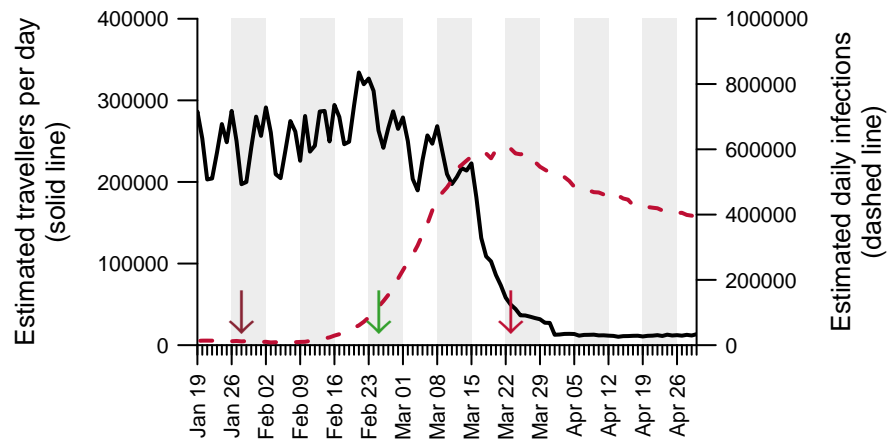


Figure 7: Estimated number of inbound travellers to the UK per day (black) and estimated number of infectious cases worldwide (dashed red, 7-day rolling average). Arrows here shows from left to right the first self-isolation advice for returning travellers from China, Italy, and the start of the UK national lockdown.

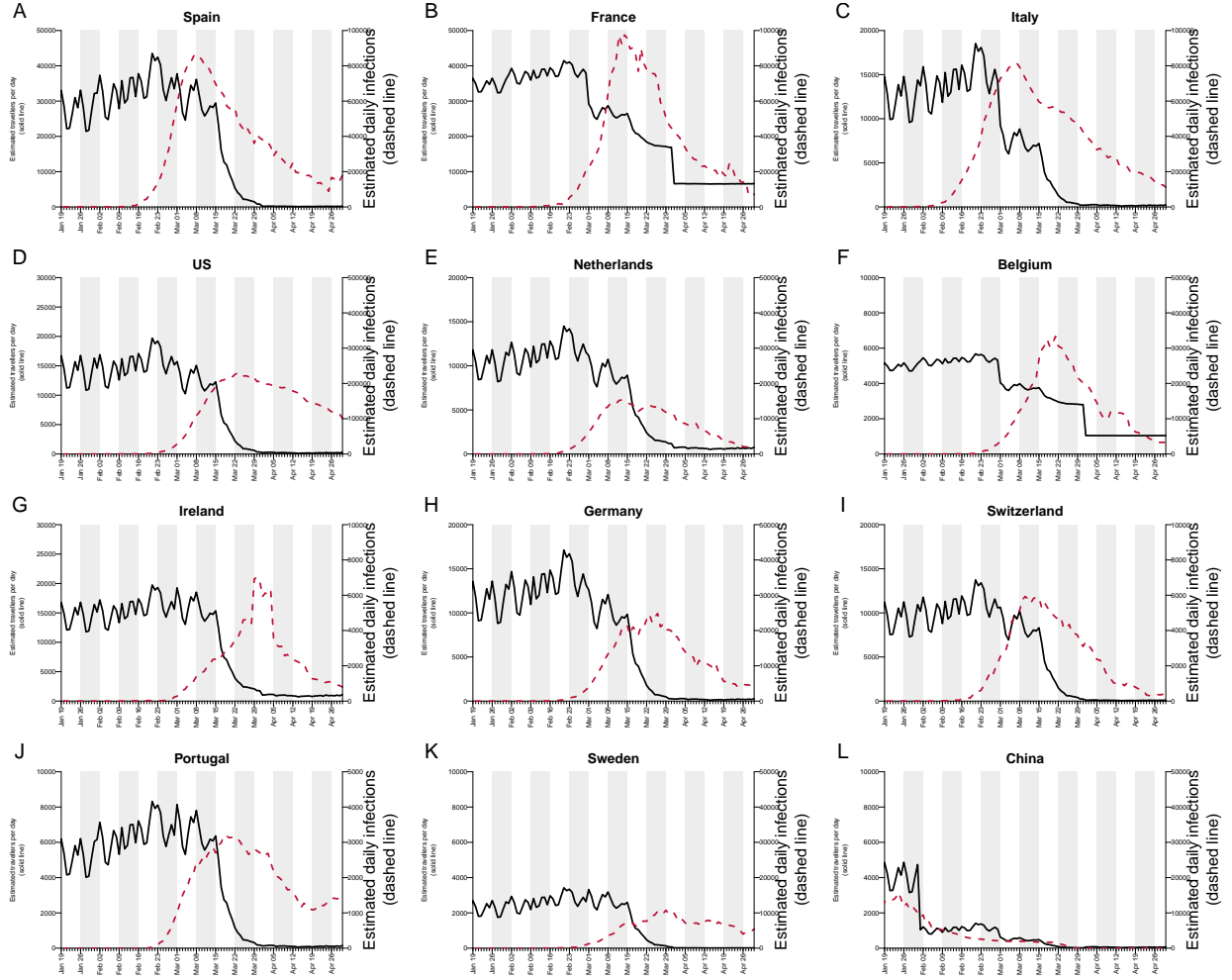


Figure 8: Estimated numbers of inbound travellers to the UK per day, and estimated number of new infections per day, for the 12 countries we estimate to have been responsible for the most importations to the UK (see Table 3). We estimate that these 12 countries contributed 97.6% of importations to the UK.

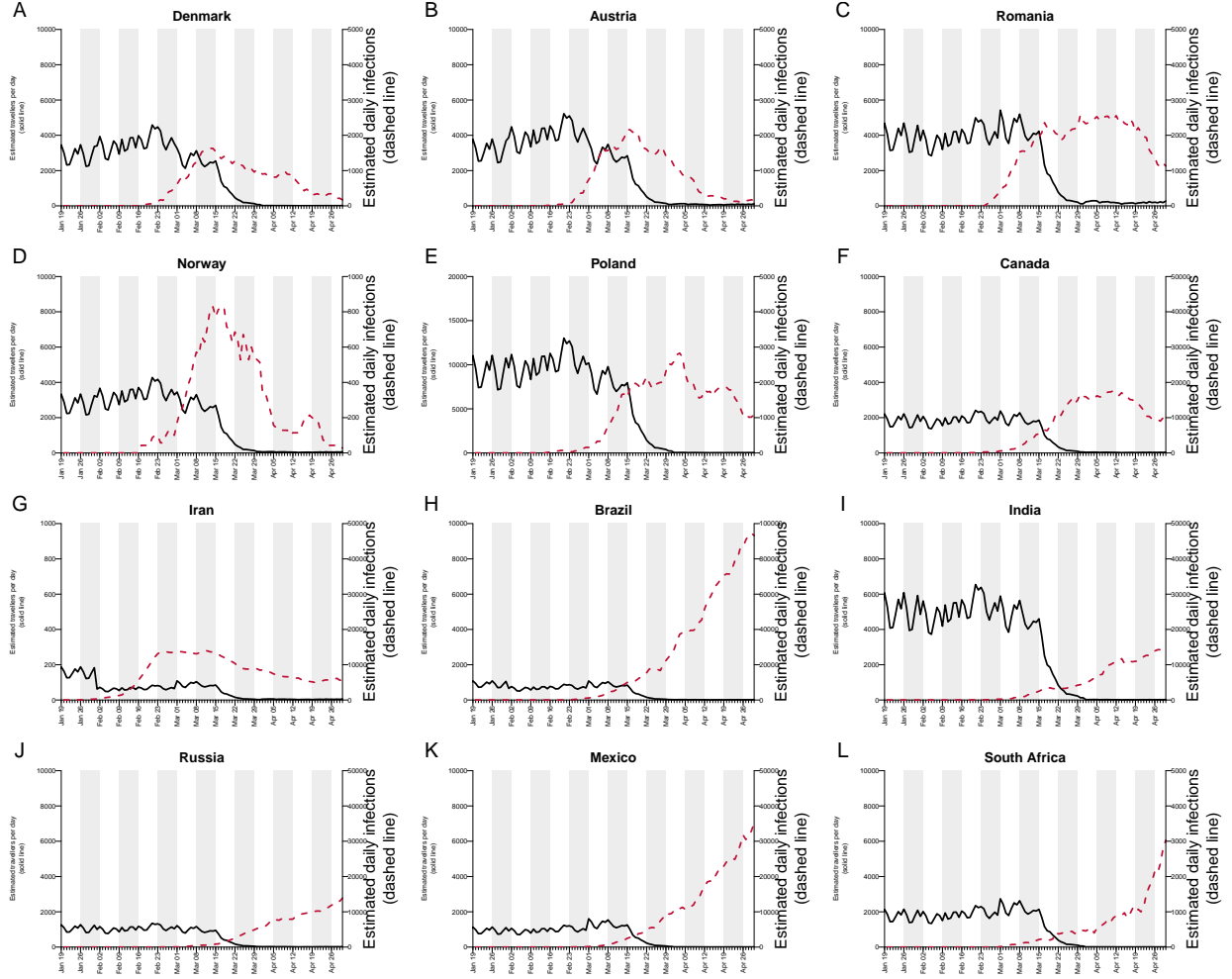


Figure 9: Estimated numbers of inbound travellers to the UK per day, and estimated number of new infections per day, for a range of countries. (A-F) shows the 6 countries that made the largest contribution to importations after the 12 in the previous figure (see Table 3). Together these 6 countries contributed 1.19% of estimated importations to the UK. (G-L) 6 countries with large epidemics that did not contribute many importations to the UK, either because of low numbers of inbound travellers or because their epidemics started later. These 6 countries combined contributed less than 0.5% of importations into the UK.

4 EII curves

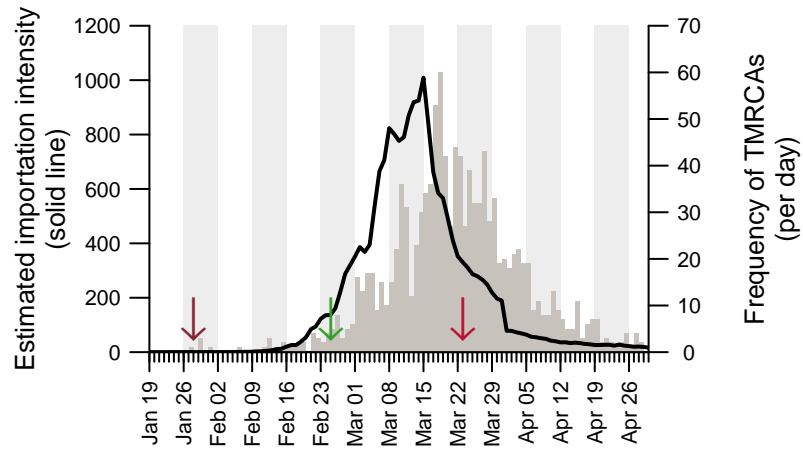


Figure 10: Estimated importation intensity (EII) curve (black) and the histogram of lineage TMRCAs (grey). Arrows here shows from left to right the first self-isolation advice for returning travellers from China, Italy, and the start of the UK national lockdown.

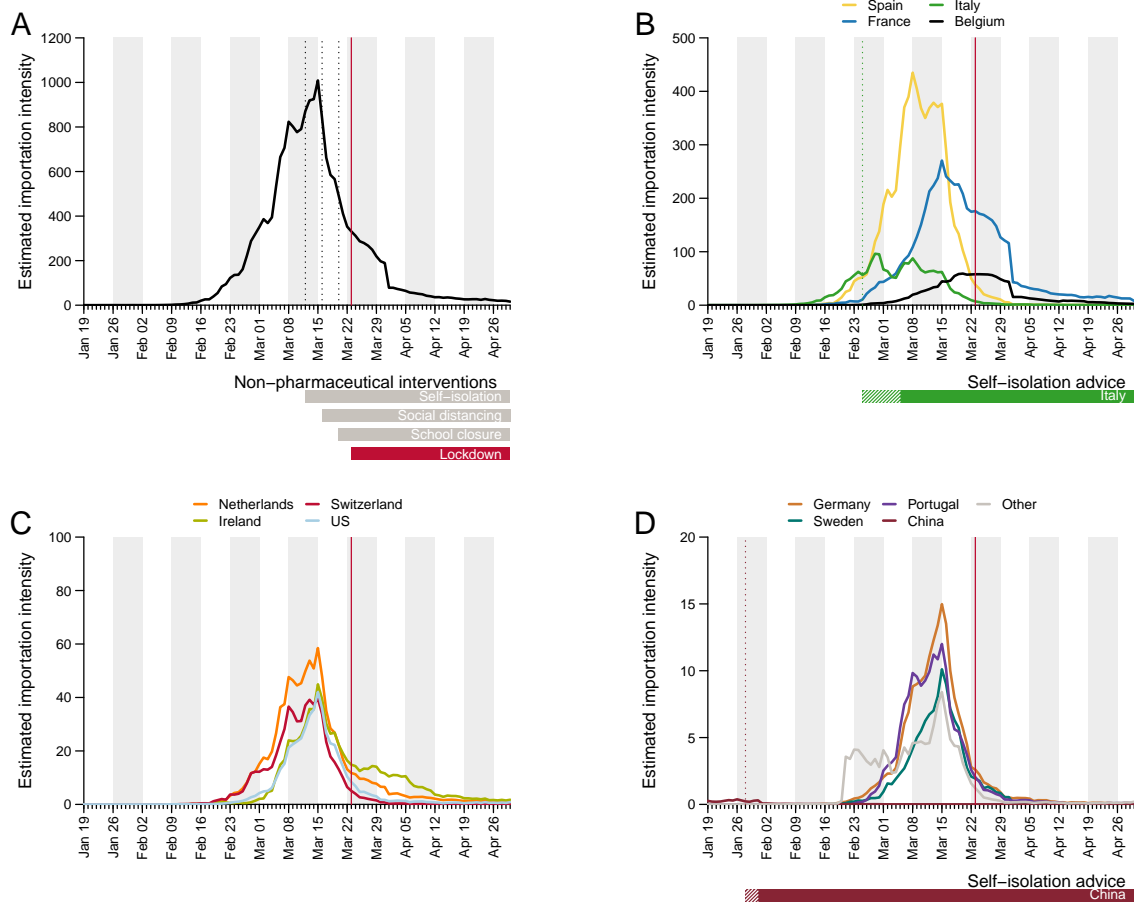


Figure 11: Estimated importation intensity (EII) curves for the 12 countries estimated to have contributed the most importations to the UK epidemic (see Table 3). Panel A shows the EII for all countries. The red arrows indicate the start of the UK lockdown.

5 Lineage importation distribution (shifted TMRCA distribution)

- Built MCC tree from 2000 posterior trees and used a threshold of 0.5 posterior probability to identify internal nodes in the UK (and identify transmission lineages)
- Dataset contains 1179 UK transmission lineages (2 or more sequences), comprising 24531 sequences from the UK, as well as a further 1650 singletons.
- Mean and SD of the importation (shifted TMRCA) distribution: 2020-03-13 \pm 13.706 days (singletons excluded).
- Median and interquartile range of the importation (shifted TMRCA) distribution 2020-03-13 [2020-03-06, 2020-03-20] (singletons excluded).
- 80% of importations fall in [2020-02-27, 2020-03-30].

Table 1: Estimated importation lags for UK transmission lineages of different sizes. Importation lag is the waiting time between importation date and the TMRCA of the sampled genomes in the transmission lineage. Detection lag is the waiting time from the importation date to the sampling time of the oldest (first) sampled genome in the transmission lineage.

Lineages of size	No. of lineages	Importation lag (mean \pm SD)	Importation lag (median and IQR)	Detection lag (mean \pm SD)	Detection lag (median and IQR)
All	1179	8.22 \pm 5.21	7.95 [3.35-15.18]	14.13 \pm 5.61	14 [10-18]
2 to 10	880	10.37 \pm 4.24	10.36 [6.5-15.18]	15.49 \pm 5	16 [12-18]
11 to 100	261	2.07 \pm 0.74	2.03 [1.41-2.65]	9.96 \pm 4.92	9 [6-13]
101 to 1000	36	0.87 \pm 0.08	0.86 [0.81-0.93]	11.08 \pm 8.03	8.5 [5.75-15]
Bigger than 1000	2	0.74 \pm 0	0.74 [0.74-0.74]	12.5 \pm 2.12	12.5 [11.75-13.25]

Table 2: 3. Estimated importation and detection lags for UK transmission lineages ordered by importation date and aggregated by epi-week. Importation lag is the waiting time between importation date and the TMRCA of the sampled genomes in the transmission lineage. Detection lag is the waiting time from the importation date to the sampling time of the oldest (first) sampled genome in the transmission lineage. All statistics show means and standard deviations computed from the MCC trees.

Week starting	Epi-week	Estimated no. of importations	Lineage sizes (median and IQR)	Importation lag (mean \pm SD)	Detection lag (mean \pm SD)
Jan 05	2	0	-	-	-
Jan 12	3	0	-	-	-
Jan 19	4	0	-	-	-
Jan 26	5	6	335.5 [140.75-742.5]	2.42 ± 3.89	20.83 ± 11.81
Feb 02	6	2	73.5 [37.75-109.25]	8.05 ± 10.08	20 ± 2.83
Feb 09	7	14	4 [2-35.5]	8.72 ± 6.81	19.07 ± 4.41
Feb 16	8	45	3 [2-12]	8.92 ± 5.8	14.27 ± 5.28
Feb 23	9	80	5.5 [2-19.5]	7.82 ± 5.85	13.43 ± 6.46
Mar 01	10	206	3 [2-10.75]	9.07 ± 5.64	14.34 ± 6.65
Mar 08	11	335	4 [2-13]	7.87 ± 5.18	14.11 ± 5.39
Mar 15	12	235	5 [2-10]	8.14 ± 4.91	13.54 ± 4.88
Mar 22	13	120	5 [3-9]	7.78 ± 4.67	13.47 ± 4.78
Mar 29	14	71	4 [2-7]	8.92 ± 4.62	14.77 ± 5.35
Apr 05	15	31	4 [3-8]	7.92 ± 4.09	15.06 ± 5.83
Apr 12	16	15	3 [2.5-5]	9.38 ± 4.67	15.73 ± 5.01
Apr 19	17	10	5 [2.25-14.75]	7.9 ± 5.74	13.9 ± 3.84
Apr 26	18	3	7 [5-10]	6.05 ± 3.85	15.33 ± 3.06
May 03	19	1	21	2.1	16
May 10	20	1	6	5.54	15
May 17	21	3	4 [3.5-6.5]	7.41 ± 3.25	14.67 ± 2.89
May 24	22	1	2	15.18	19
May 31	23	0	-	-	-
Jun 07	24	0	-	-	-
Jun 14	25	0	-	-	-

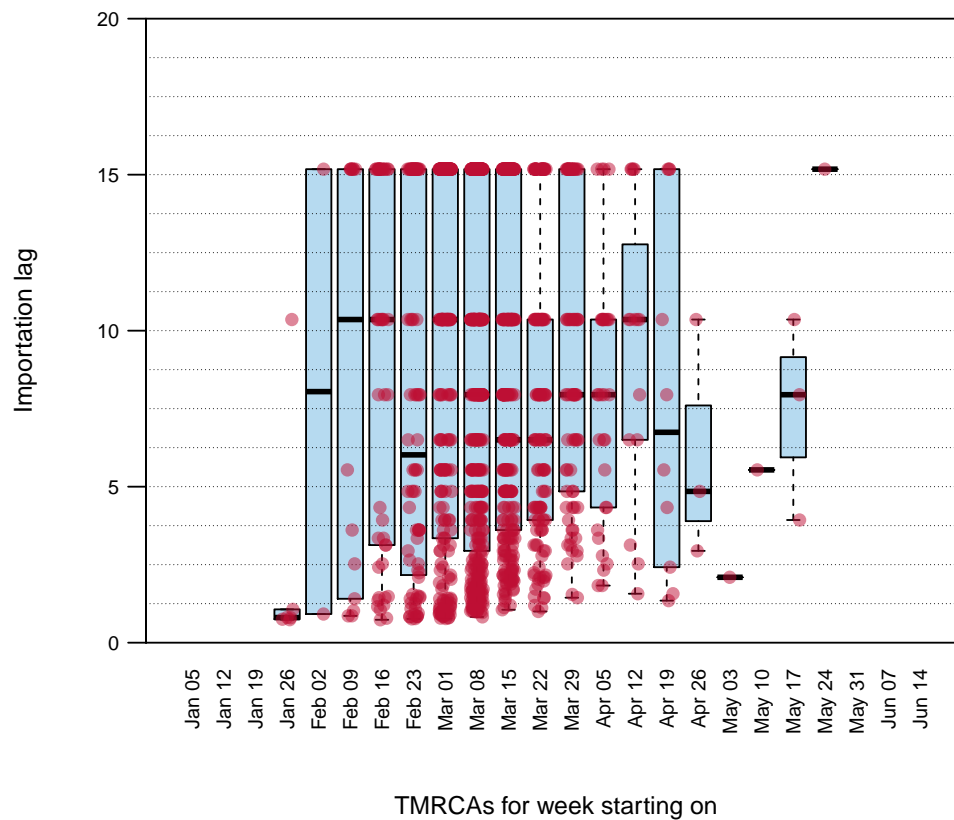


Figure 12: Boxplots of the estimated importation lags for UK transmission lineages ordered by importation date and aggregated by epi-week.

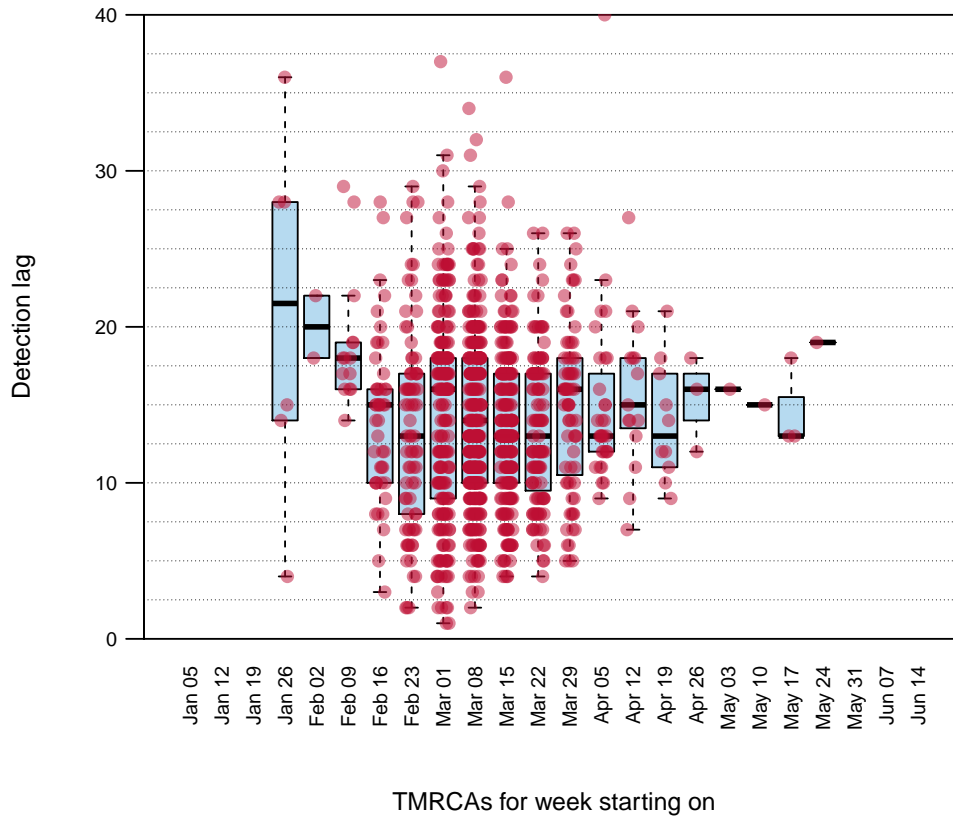


Figure 13: Boxplots of the estimated detection lags for UK transmission lineages ordered by importation date and aggregated by epi-week.

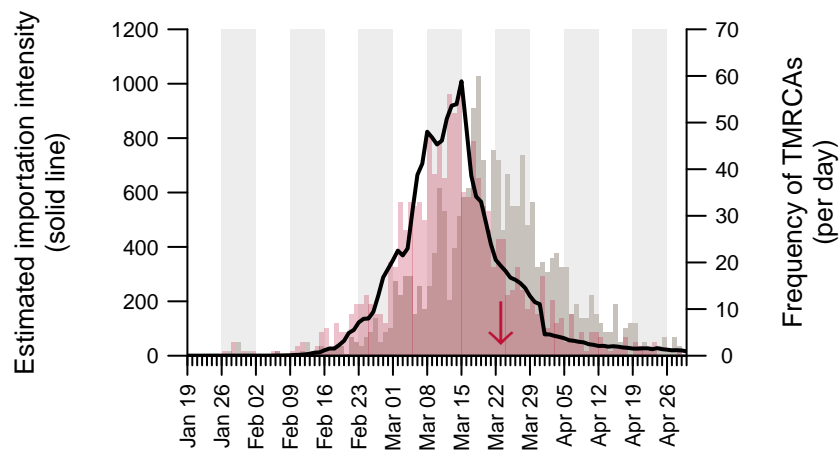


Figure 14: Estimated importation intensity (EII) curve (black) and the histogram of lineage TMRCAs (grey) and shifted TMRCAs representing importations (red). Arrow here shows the start of the UK lockdown.

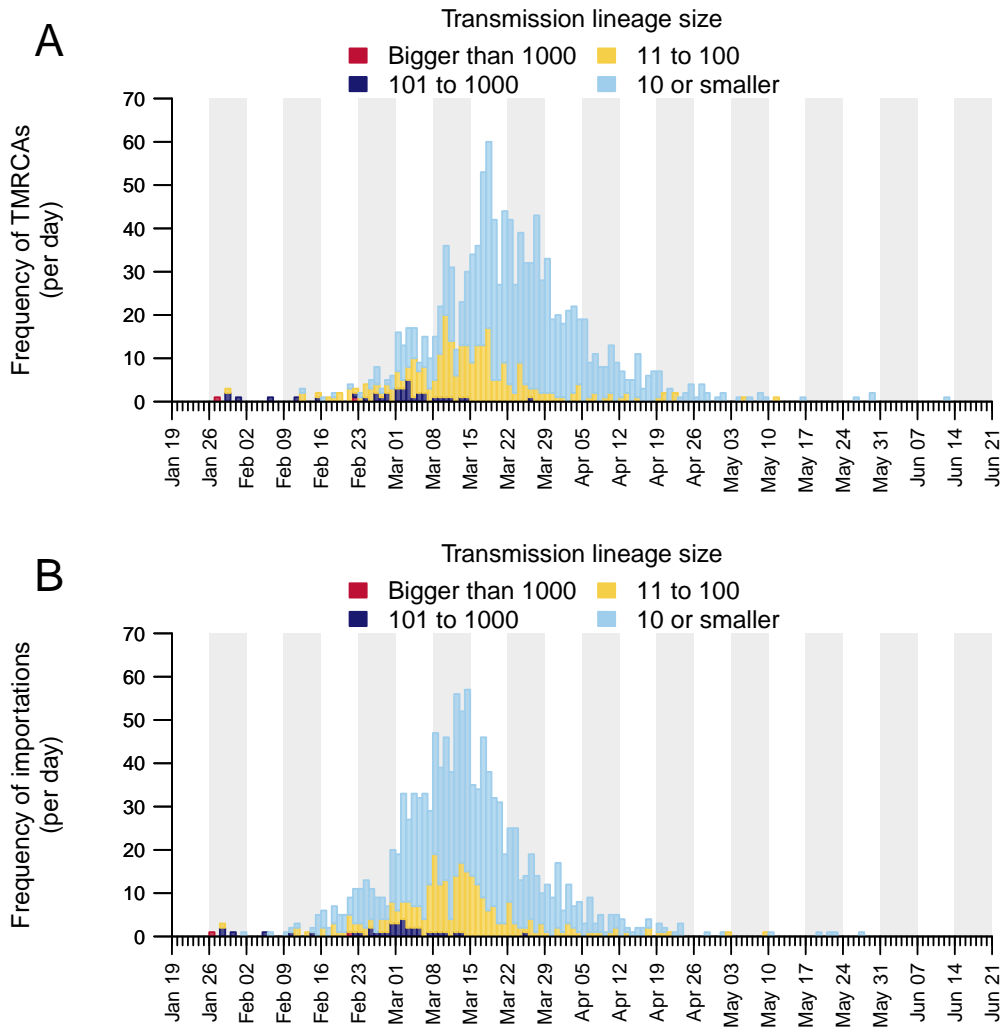


Figure 15: (A) Histogram of lineage TMRCAs, coloured by lineage size. (B) Histogram of lineage importations, coloured by lineage size.

6 Import breakdown

##

##

Numbers of importations across 2000 posterior trees: : 2413.805 sec elapsed

Table 3: Number of observed importations in our dataset and the percentage of the total that can be attributed to the 40 countries inferred to be sources for the most importations on the MCC tree and across a set of 1989 posterior trees (the optimisation procedure failed to converge on 11 posterior trees).

	Observed importations (MCC tree)	Observed importations (median, 95% HPD)	Percentage (MCC tree)	Percentage (median, 95% HPD)
Spain	387.12	399.47 [375.89-424.78]	33.066	32.963 [32.383-33.563]
France	334.04	344.25 [321.8-367.25]	28.532	28.433 [27.942-28.908]
Italy	140.83	149.12 [139.25-158.69]	12.029	12.327 [11.673-12.943]
Belgium	84.88	87.61 [80.99-93.28]	7.250	7.239 [7.029-7.453]
Netherlands	55.08	56.84 [53.25-60.29]	4.705	4.694 [4.642-4.752]
Ireland	45.30	46.63 [43.3-49.65]	3.869	3.848 [3.762-3.943]
Switzerland	34.91	36.08 [33.85-38.29]	2.982	2.978 [2.924-3.024]
US	29.16	30.21 [28.08-32.37]	2.491	2.494 [2.44-2.559]
Germany	10.85	11.2 [10.37-11.93]	0.927	0.924 [0.905-0.947]
Portugal	9.56	9.87 [9.17-10.54]	0.817	0.815 [0.796-0.833]
Sweden	6.71	6.87 [6.34-7.35]	0.573	0.567 [0.552-0.585]
China	4.64	3.97 [2-6.52]	0.397	0.328 [0.155-0.532]
Denmark	3.84	3.94 [3.68-4.2]	0.328	0.325 [0.318-0.332]
Austria	3.48	3.59 [3.34-3.82]	0.297	0.297 [0.29-0.303]
Romania	2.24	2.33 [2.19-2.51]	0.191	0.193 [0.188-0.198]
Norway	1.95	2.02 [1.88-2.16]	0.167	0.167 [0.163-0.171]
Poland	1.28	1.32 [1.22-1.41]	0.109	0.109 [0.106-0.112]
Canada	1.08	1.14 [1.06-1.22]	0.093	0.094 [0.092-0.096]
Turkey	0.96	0.99 [0.92-1.06]	0.082	0.082 [0.08-0.084]
Hungary	0.95	1 [0.94-1.07]	0.081	0.082 [0.08-0.084]
Czechia	0.65	0.67 [0.62-0.72]	0.056	0.055 [0.053-0.057]
Greece	0.55	0.57 [0.54-0.61]	0.047	0.047 [0.046-0.048]
United Arab Emirates	0.30	0.32 [0.3-0.34]	0.026	0.026 [0.026-0.027]
Israel	0.27	0.28 [0.25-0.29]	0.023	0.023 [0.022-0.023]
Finland	0.25	0.26 [0.24-0.28]	0.022	0.022 [0.021-0.022]
Iran	0.22	0.24 [0.22-0.26]	0.019	0.019 [0.018-0.021]
South Korea	0.20	0.25 [0.18-0.33]	0.017	0.02 [0.014-0.027]
Morocco	0.18	0.18 [0.17-0.19]	0.015	0.015 [0.014-0.015]
Brazil	0.17	0.19 [0.17-0.2]	0.015	0.015 [0.015-0.016]
Dominican Republic	0.16	0.16 [0.15-0.17]	0.013	0.013 [0.013-0.014]
Mexico	0.08	0.09 [0.08-0.09]	0.007	0.007 [0.007-0.007]
Serbia	0.07	0.08 [0.07-0.08]	0.006	0.006 [0.006-0.007]
Japan	0.07	0.08 [0.06-0.09]	0.006	0.006 [0.005-0.008]

	Observed importations (MCC tree)	Observed importations (median, 95% HPD)	Percentage (MCC tree)	Percentage (median, 95% HPD)
Egypt	0.06	0.06 [0.05-0.06]	0.005	0.005 [0.005-0.005]
Malaysia	0.05	0.05 [0.05-0.06]	0.004	0.004 [0.004-0.005]
Pakistan	0.05	0.06 [0.05-0.06]	0.004	0.005 [0.004-0.005]
Moldova	0.04	0.04 [0.04-0.05]	0.004	0.004 [0.003-0.004]
Philippines	0.04	0.04 [0.04-0.04]	0.003	0.003 [0.003-0.003]
Russia	0.03	0.04 [0.03-0.04]	0.003	0.003 [0.003-0.003]
Ecuador	0.03	0.03 [0.03-0.03]	0.003	0.003 [0.002-0.003]
Other	8.39	8.73 [8.2-9.37]	0.717	0.721 [0.693-0.749]

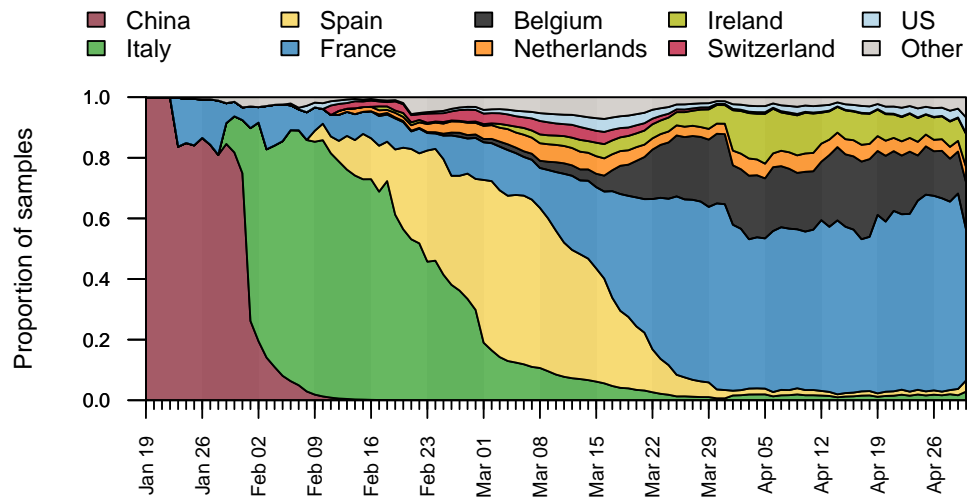


Figure 16: The estimated proportion of importation events that are attributable to inbound travellers from each of several source countries over time.

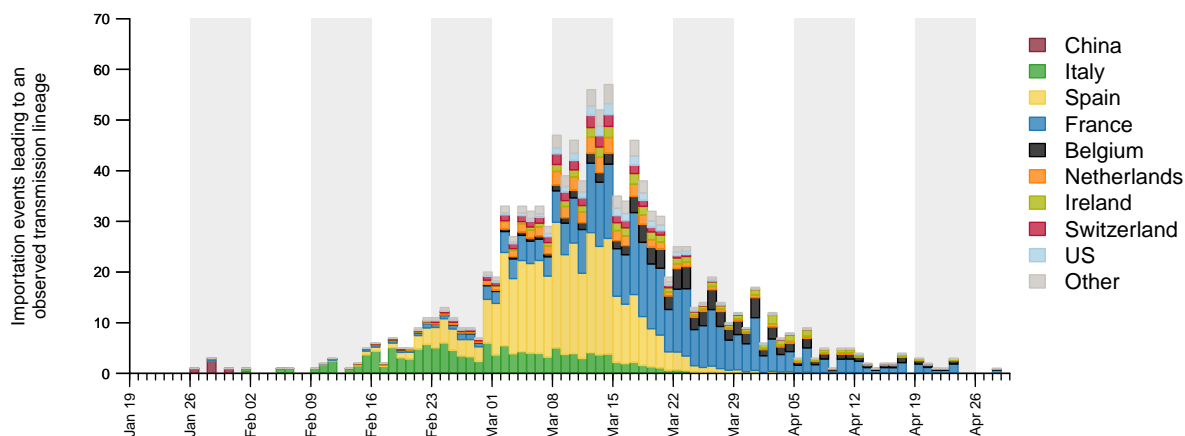


Figure 17: Estimated histogram of virus lineage importation events per day, obtained from our lag model. Colours show the proportion attributable each day to inbound travel from various countries. This assignment is statistical, i.e. we cannot ascribe a specific source location to any given lineage.

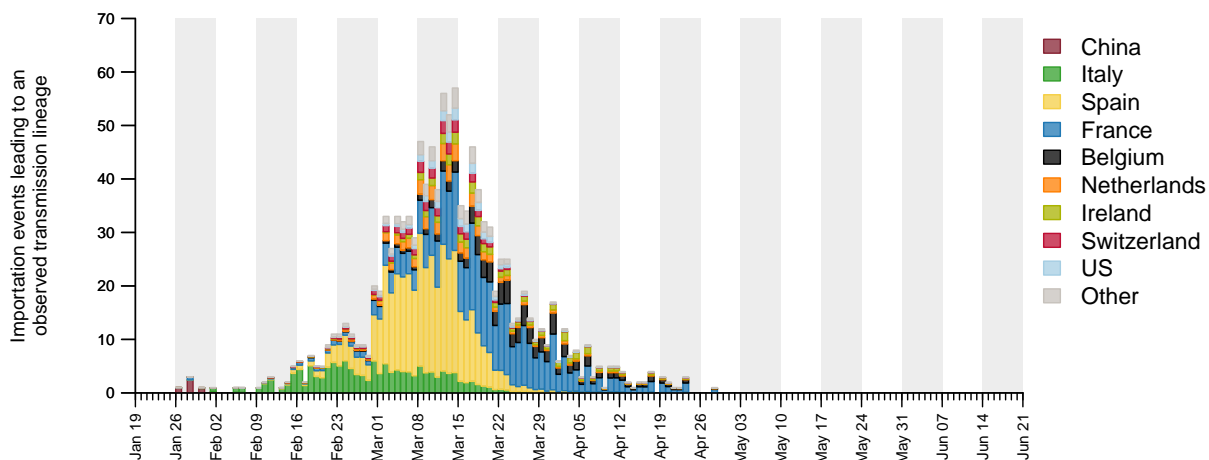


Figure 18: Estimated histogram of virus lineage importation events per day, obtained from our lag model. Colours show the proportion attributable each day to inbound travel from various countries. This assignment is statistical, i.e. we cannot ascribe a specific source location to any given lineage. **This is the same as the previous plot but the axes have been extended. Lineages with TMRCAs after 30 April are not plotted, since data on inbound travellers after 30 April are not available.**

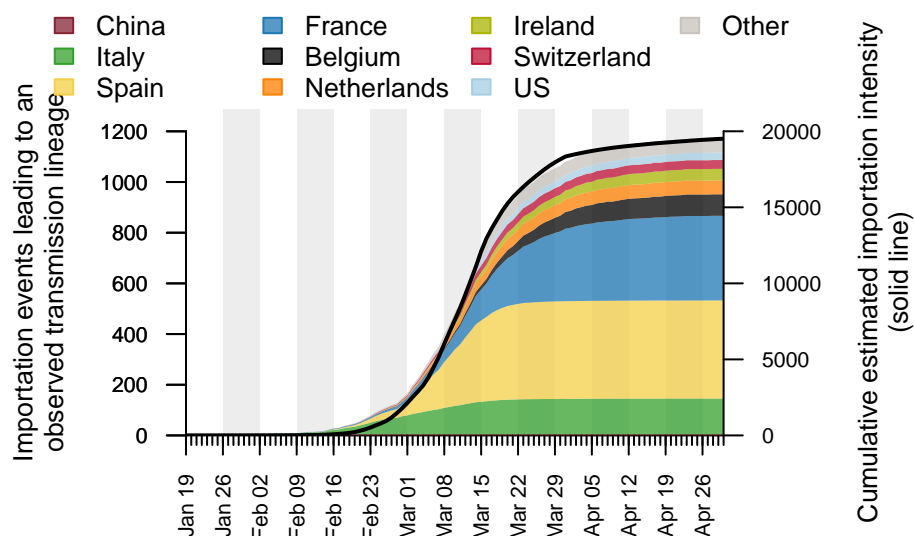


Figure 19: Estimated cumulative number of virus lineage importation events per day, obtained from our lag model. Colours show the proportion attributable each day to inbound travel from various countries. This assignment is statistical, i.e. we cannot ascribe a specific source location to any given lineage. The black line shows the cumulative global EII curve. **This plot is to check the fit of the importation distribution to the EII curve by eye.**

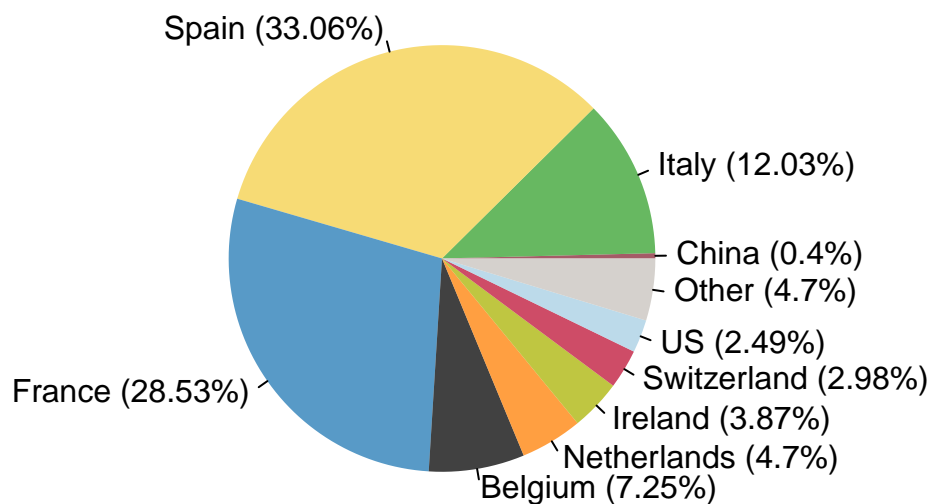


Figure 20: The estimated total fraction of importation events that are attributable to inbound travellers from each country.

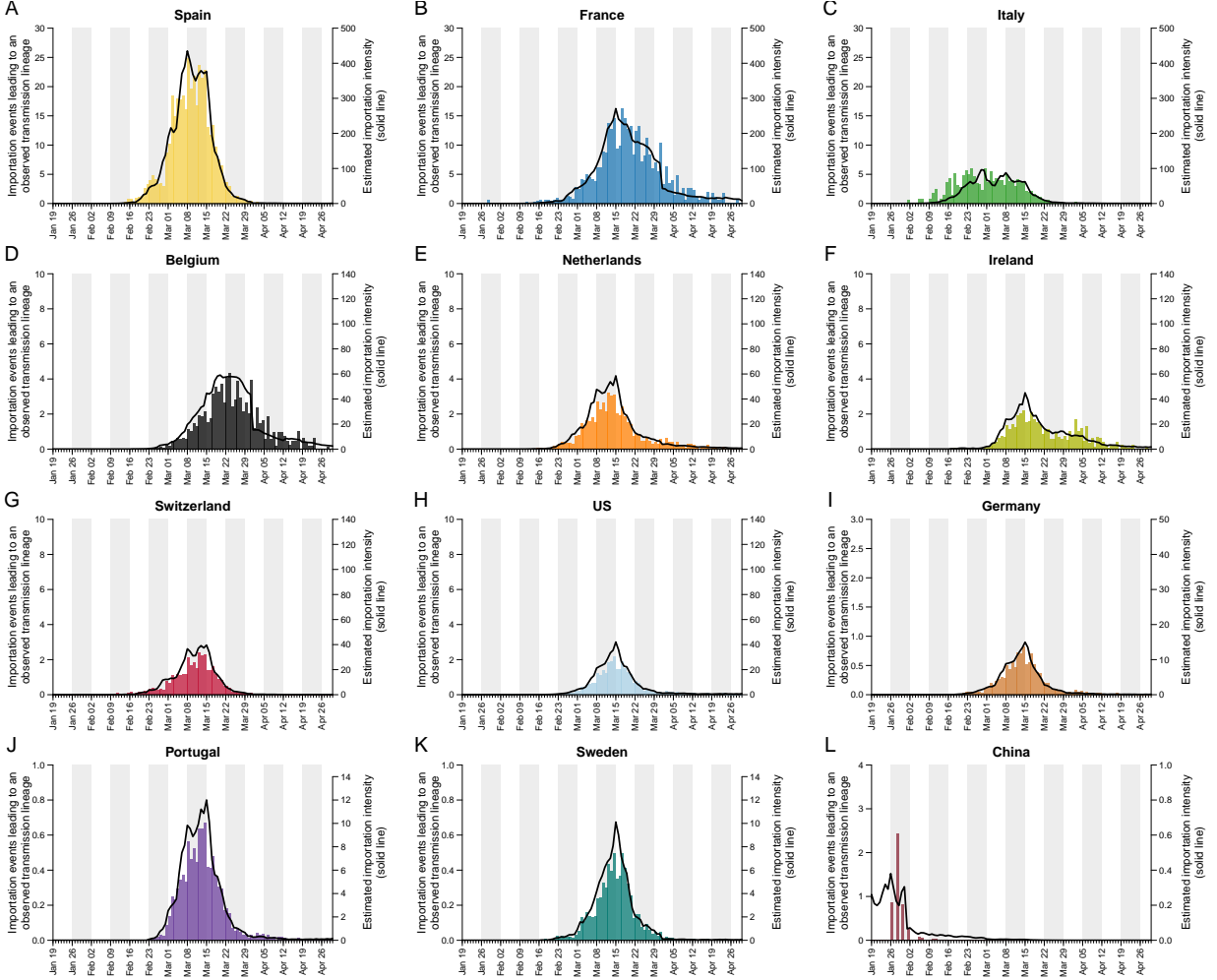


Figure 21: Estimated histogram of virus lineage importation events per day, obtained from our lag model, for the 12 countries estimated to have contributed the most importations (see Table 3). This assignment is statistical, i.e. we cannot ascribe a specific source location to any given lineage. **This plot is just a sanity check and isn't useful for drawing any conclusions.**

7 Session info

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Sierra 10.12.6
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] phytools_0.6-99 maps_3.3.0      ape_5.3          ggtree_2.1.1
## [5] ggsci_2.9      ggplot2_3.2.1  tictoc_1.0       knitr_1.29
## [9] beastio_0.3.3  gplots_3.0.1.1 plyr_1.8.5       lubridate_1.7.4
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.3          lattice_0.20-38      tidyr_1.0.0
## [4] gtools_3.8.1        assertthat_0.2.1     zeallot_0.1.0
## [7] digest_0.6.23       R6_2.4.1             backports_1.1.5
## [10] coda_0.19-3         evaluate_0.14        highr_0.8
## [13] pillar_1.4.2        rlang_0.4.2          lazyeval_0.2.2
## [16] gdata_2.18.0        phangorn_2.5.5       combinat_0.0-8
## [19] Matrix_1.2-18       rmarkdown_2.3        stringr_1.4.0
## [22] igraph_1.2.4.2       munsell_0.5.0        compiler_3.5.1
## [25] numDeriv_2016.8-1.1 xfun_0.15            pkgconfig_2.0.3
## [28] mnormt_1.5-5        htmltools_0.4.0      tidyselect_0.2.5
## [31] expm_0.999-4        tibble_2.1.3         quadprog_1.5-8
## [34] crayon_1.3.4        dplyr_0.8.3          withr_2.1.2
## [37] MASS_7.3-51.4       bitops_1.0-6         grid_3.5.1
## [40] nlme_3.1-143        jsonlite_1.6          gtable_0.3.0
## [43] lifecycle_0.1.0     magrittr_1.5          scales_1.1.0
## [46] KernSmooth_2.23-16  tidytree_0.3.1       stringi_1.4.3
## [49] scatterplot3d_0.3-41 rvcheck_0.1.7        vctrs_0.2.1
## [52] fastmatch_1.1-0     RColorBrewer_1.1-2   tools_3.5.1
## [55] treeio_1.11.2       glue_1.3.1           purrr_0.3.3
## [58] plotrix_3.7-7       parallel_3.5.1       yaml_2.2.0
## [61] colorspace_1.4-1    BiocManager_1.30.10  caTools_1.17.1.3
## [64] animation_2.6       clusterGeneration_1.3.4
```