

COVID-19 UK introductions

Transmission lineage breakdown (detailed descriptions)

Louis du Plessis

Last modified: 07 Dec 2020

Contents

1	Summary	1
1.1	Input	1
1.2	Output	1
2	Sample breakdown (daily)	2
3	Sample breakdown into lineages (weekly)	5
4	Sample breakdown into lineages (daily)	10
5	Individual transmission lineage plots	15
5.1	Biggest transmission lineages	15
5.2	Earliest transmission lineages	15
5.3	Newest transmission lineages	15
5.4	Longest periods of cryptic circulation	15
5.5	Longest sampling period	15
5.6	Longest unobserved period before reactivating	15
6	Session info	22

1 Summary

This notebook plots breakdowns of UK transmission lineages over time (using only the assignment on the MCC trees).

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`

1.2 Output

- Lineage breakdown figures and tables.

2 Sample breakdown (daily)

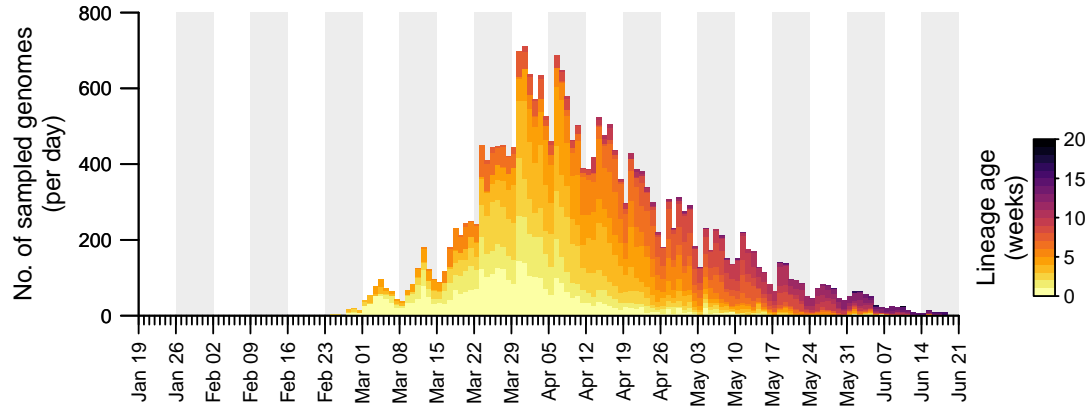


Figure 1: Number of genomes collected each day, coloured by the time since the age of the transmission lineage when the genome was collected (time from the oldest sampled genome in the lineage to the sampling time of the genome). Note that **only** genomes in transmission lineages are shown (no singletons).

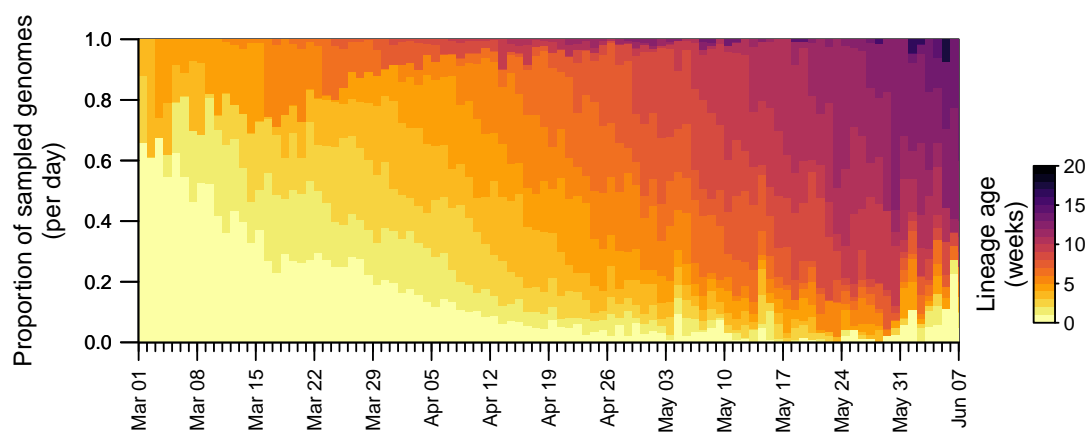


Figure 2: Proportion of genomes collected each day, coloured by the time since the age of the transmission lineage when the genome was collected (time from the oldest sampled genome in the lineage to the sampling time of the genome). Note that **only** genomes in transmission lineages are shown (no singletons).

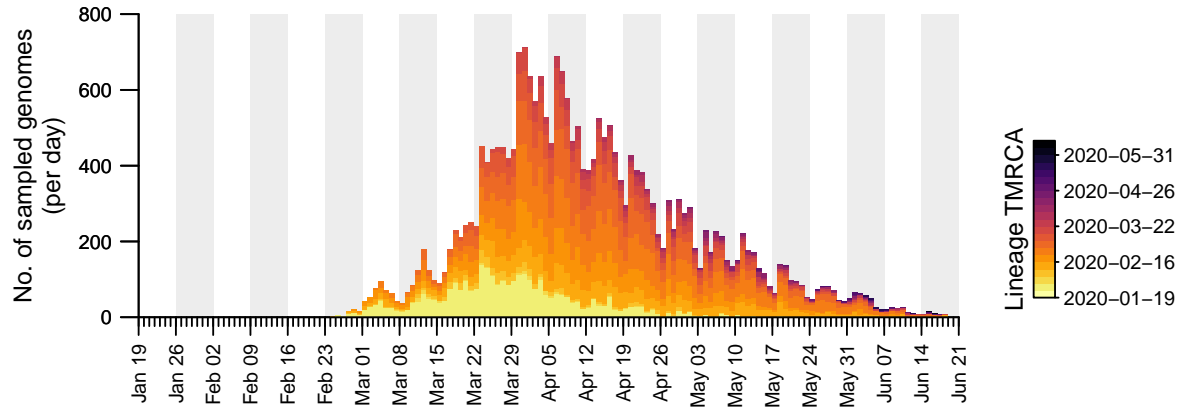


Figure 3: Number of genomes collected each day, coloured by the TMRCA of the transmission lineage. Note that **only** genomes in transmission lineages are shown (no singletons).

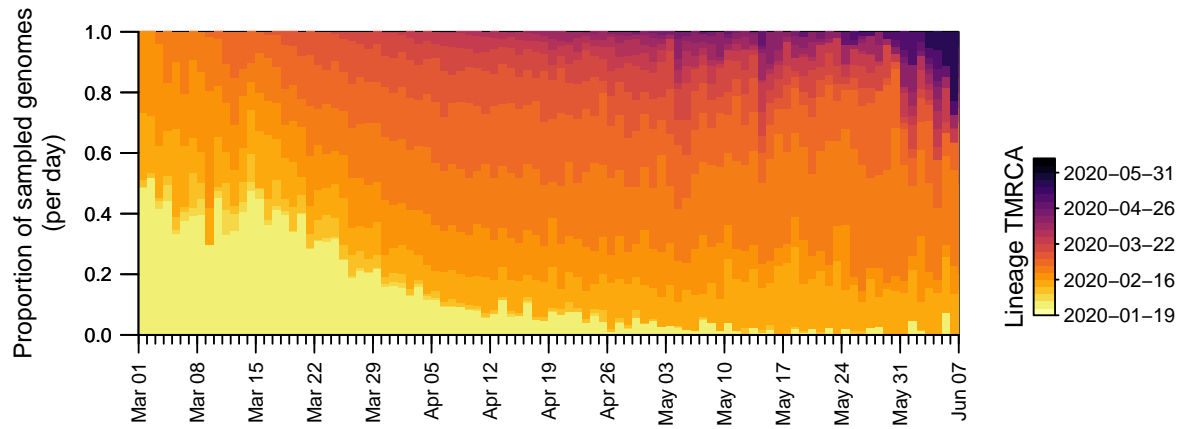


Figure 4: Proportion of genomes collected each day, coloured by the TMRCA of the transmission lineage. Note that **only** genomes in transmission lineages are shown (no singletons).

3 Sample breakdown into lineages (weekly)

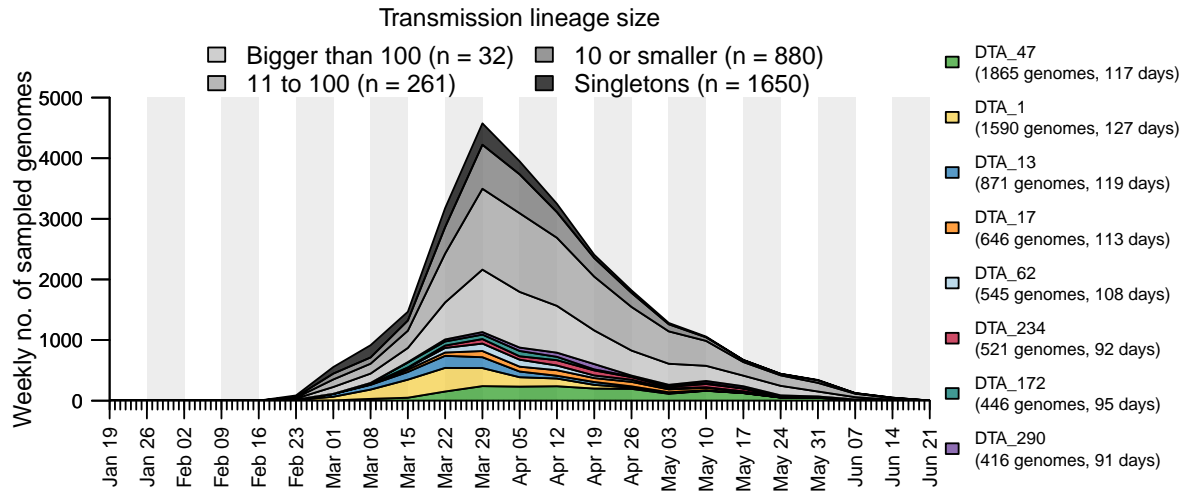


Figure 5: Lineage size breakdown of UK genomes collected each week. The 8 largest lineages are coloured.

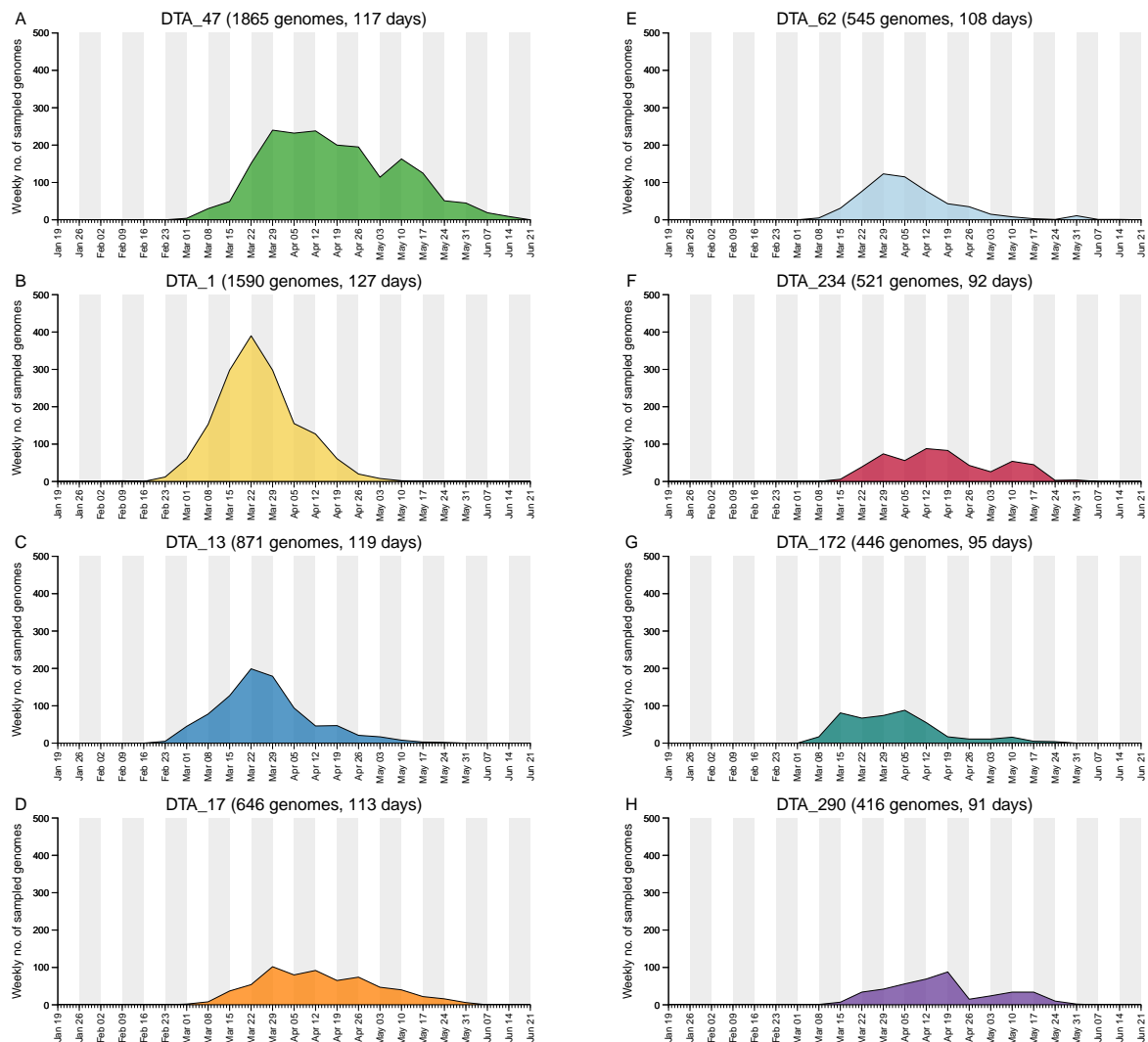


Figure 6: The weekly sampling frequency of the 8 largest UK transmission lineages.

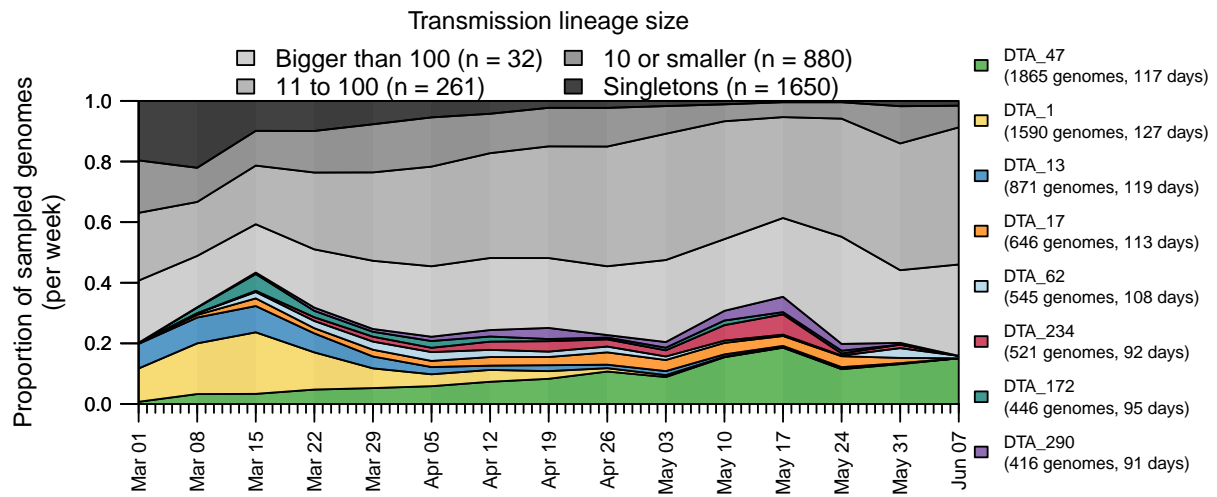


Figure 7: Lineage size breakdown of UK genomes collected each week. The 8 largest lineages are coloured.

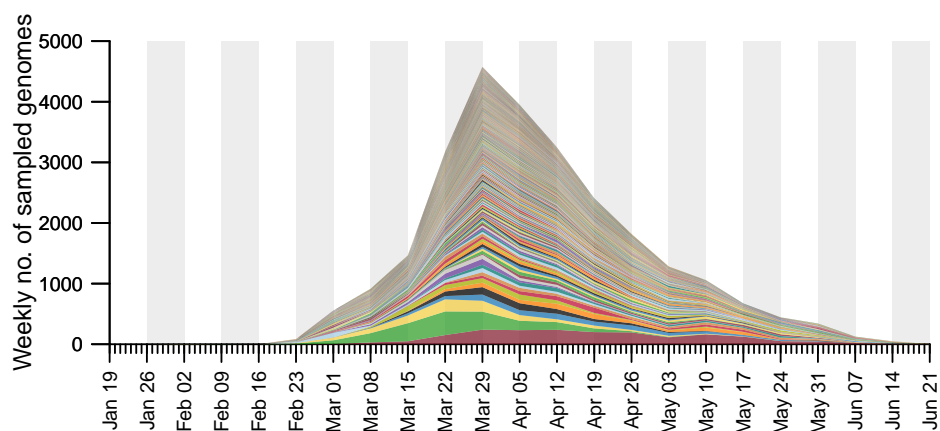


Figure 8: Lineage size breakdown of UK genomes collected each week.

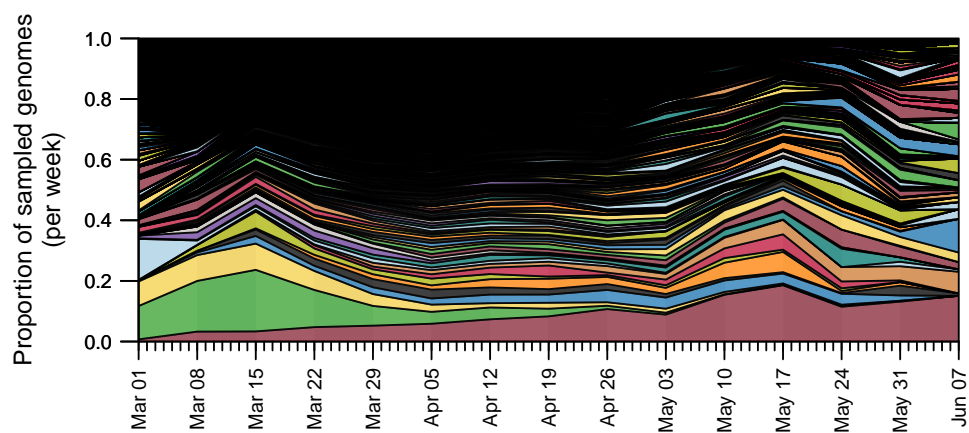


Figure 9: Lineage size breakdown of UK genomes collected each week.

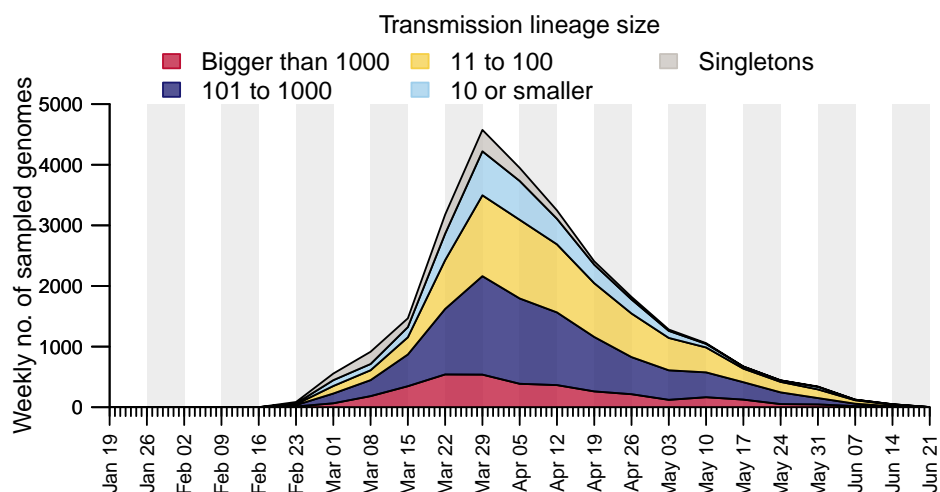


Figure 10: Lineage size breakdown of UK genomes collected each week.

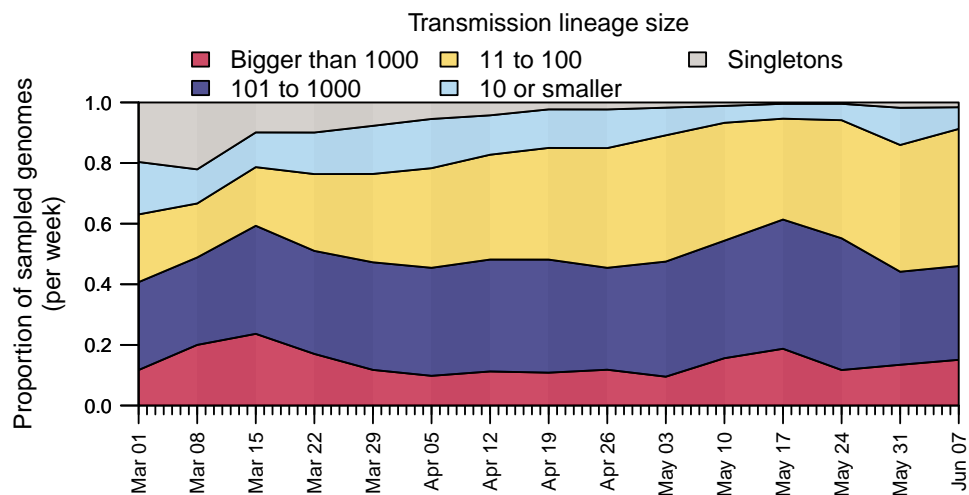


Figure 11: Lineage size breakdown of UK genomes collected each week.

4 Sample breakdown into lineages (daily)

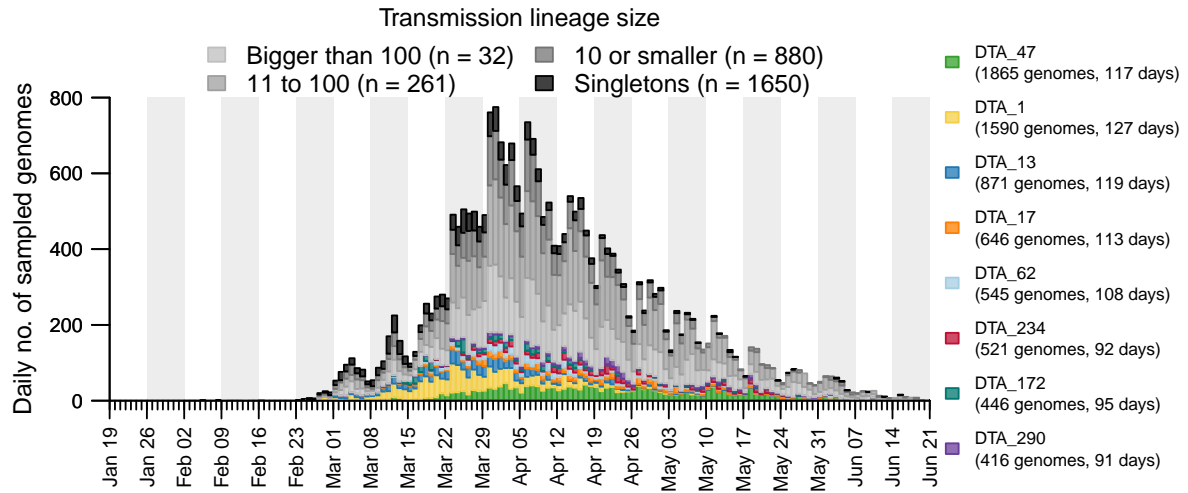


Figure 12: Lineage size breakdown of UK genomes collected each day. The 8 largest lineages are coloured.

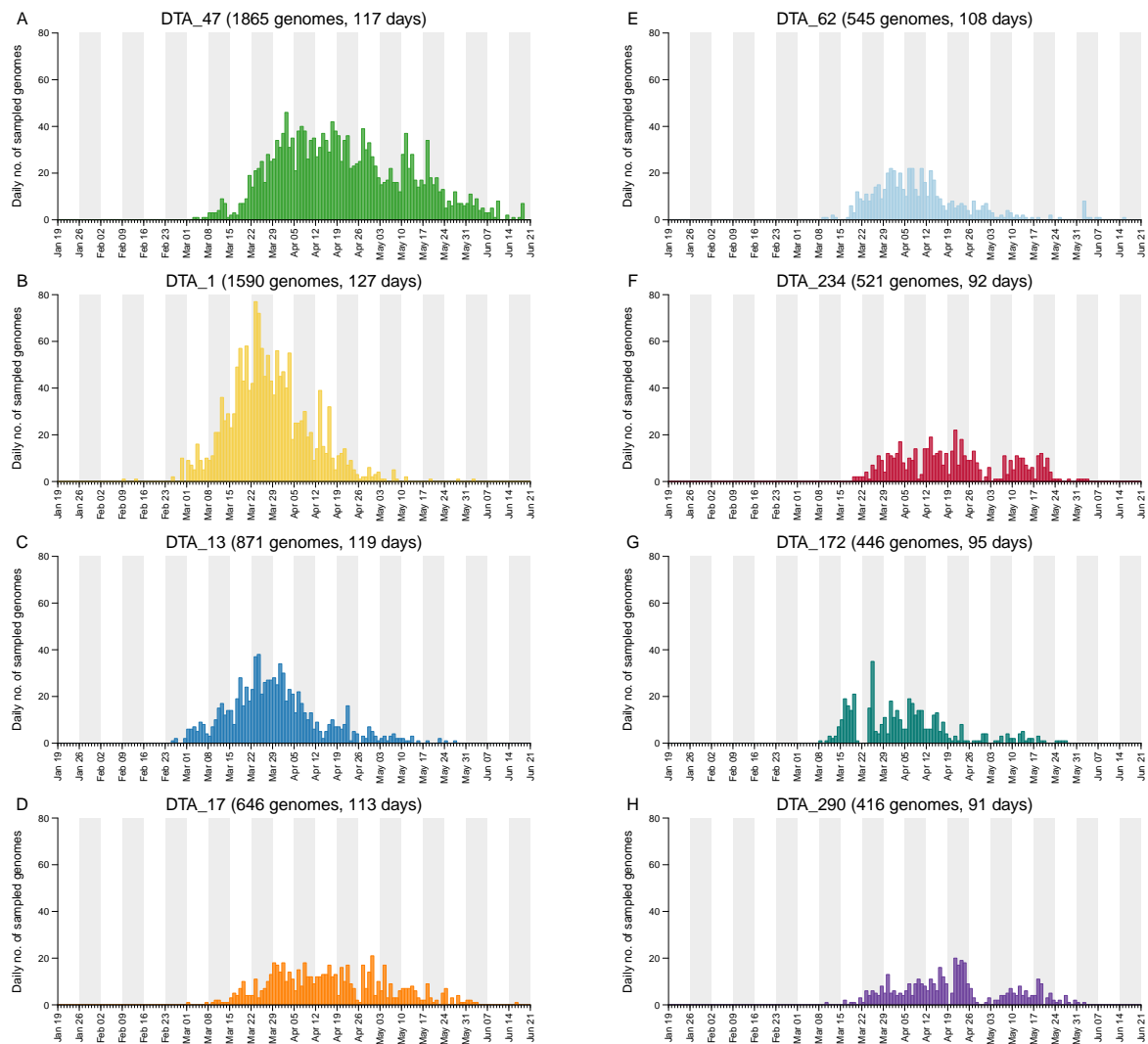


Figure 13: The daily sampling frequency of the 8 largest UK transmission lineages.

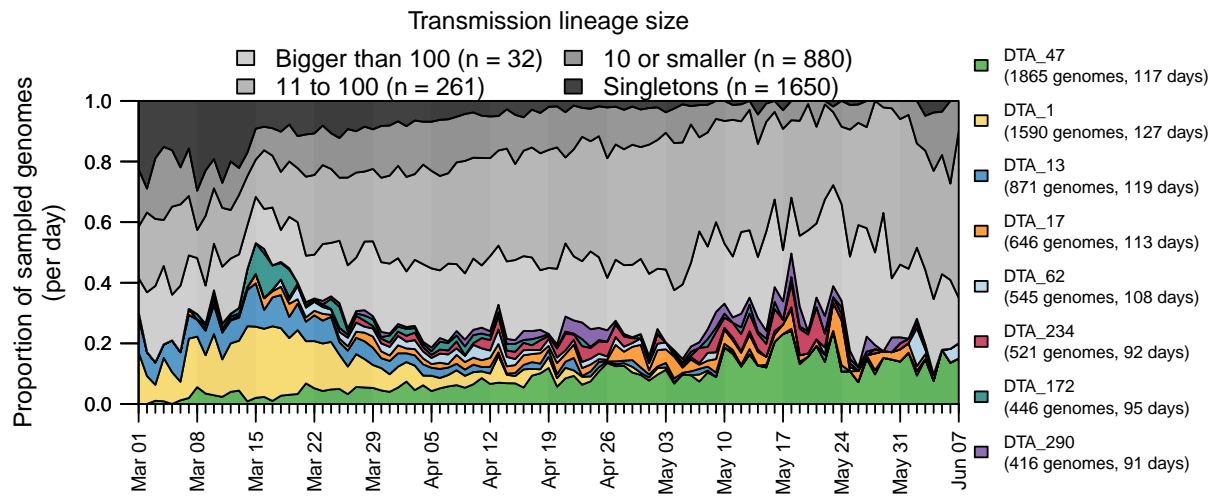


Figure 14: Lineage size breakdown of UK genomes collected each day. The 8 largest lineages are coloured.

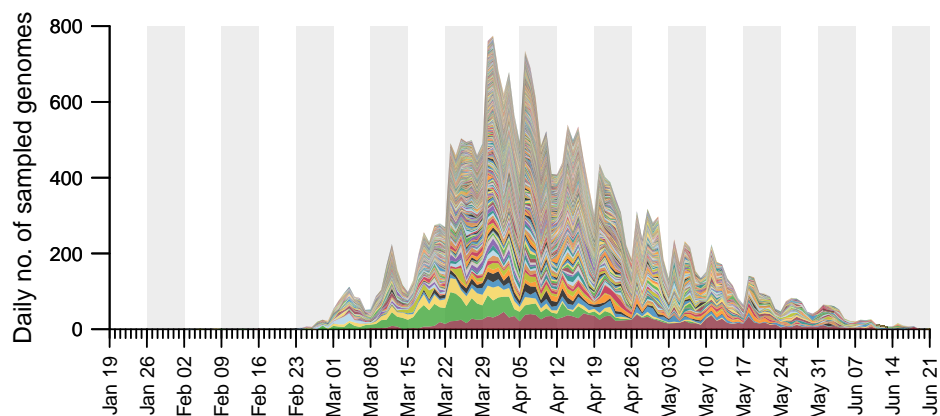


Figure 15: Lineage size breakdown of UK genomes collected each day.

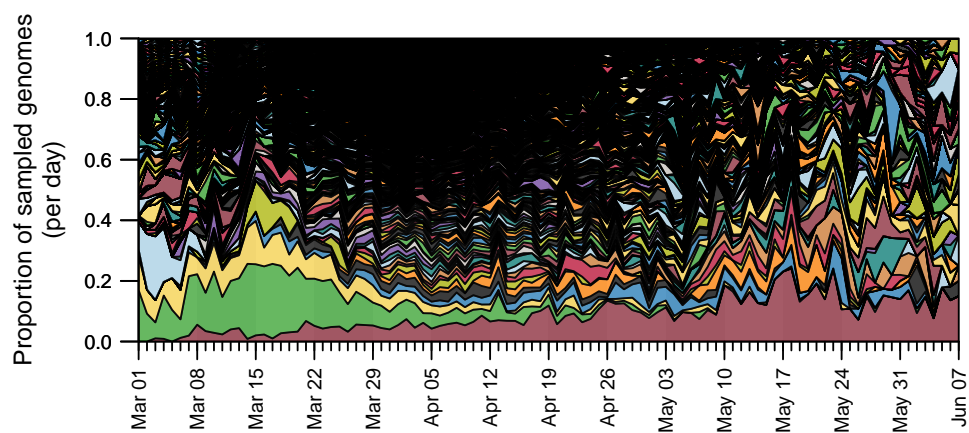


Figure 16: Lineage size breakdown of UK genomes collected each day.

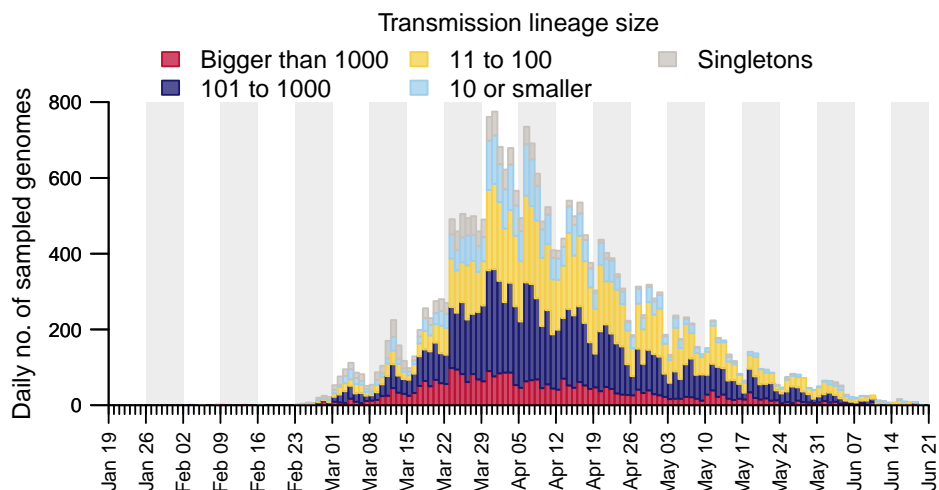


Figure 17: Lineage size breakdown of UK genomes collected each day.

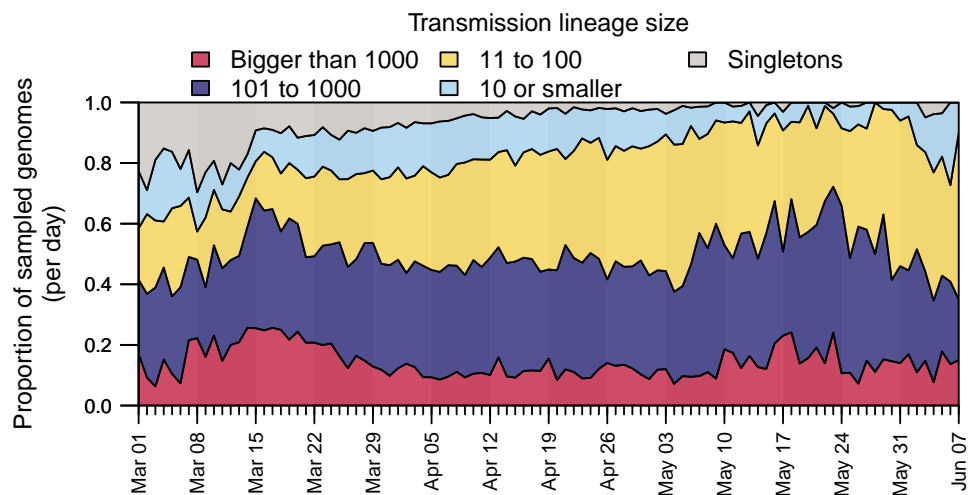


Figure 18: Lineage size breakdown of UK genomes collected each day.

5 Individual transmission lineage plots

5.1 Biggest transmission lineages

5.2 Earliest transmission lineages

5.3 Newest transmission lineages

5.4 Longest periods of cryptic circulation

5.5 Longest sampling period

5.6 Longest unobserved period before reactivating

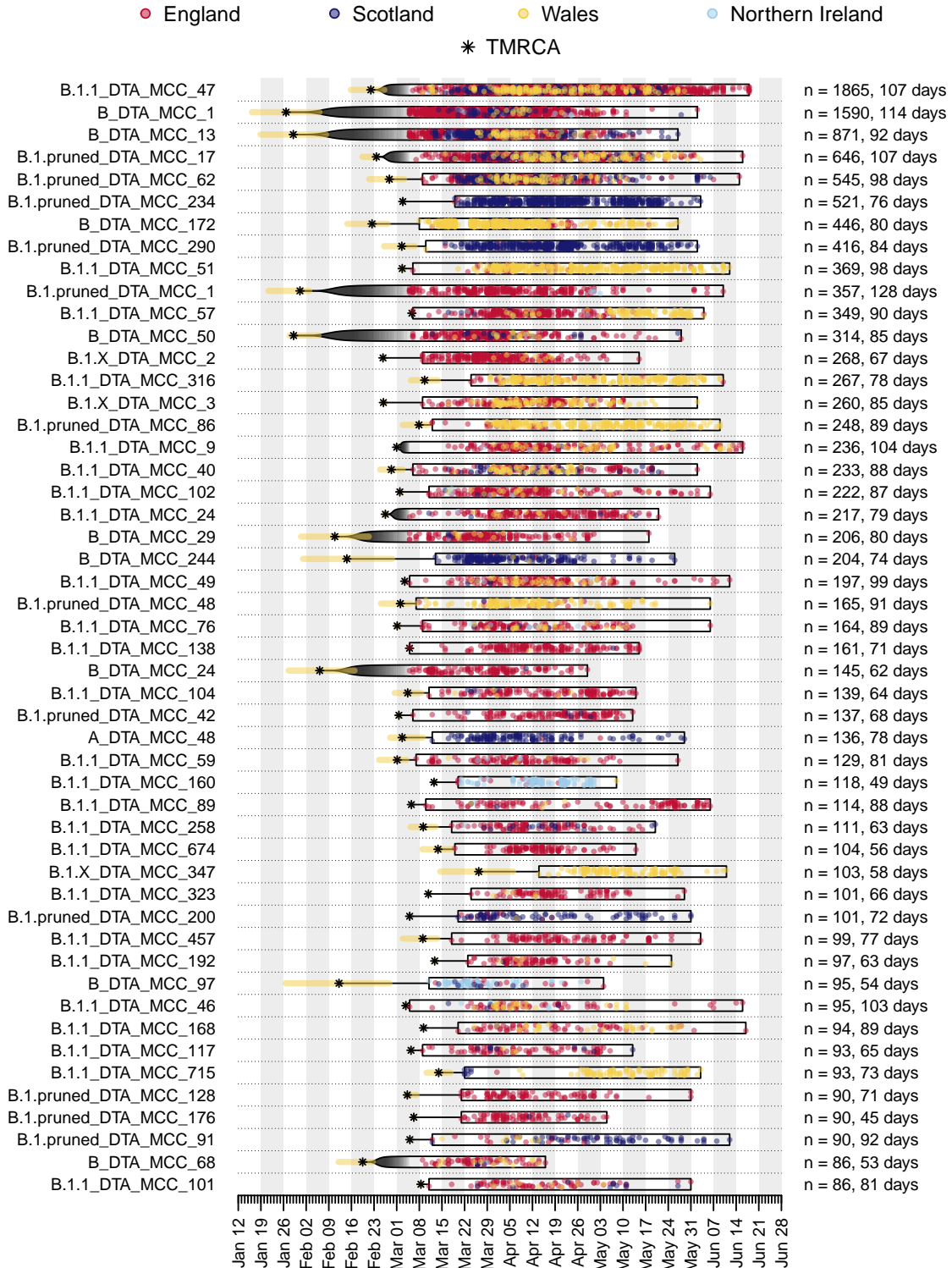


Figure 19: Illustration of the time course of the 50 largest UK transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

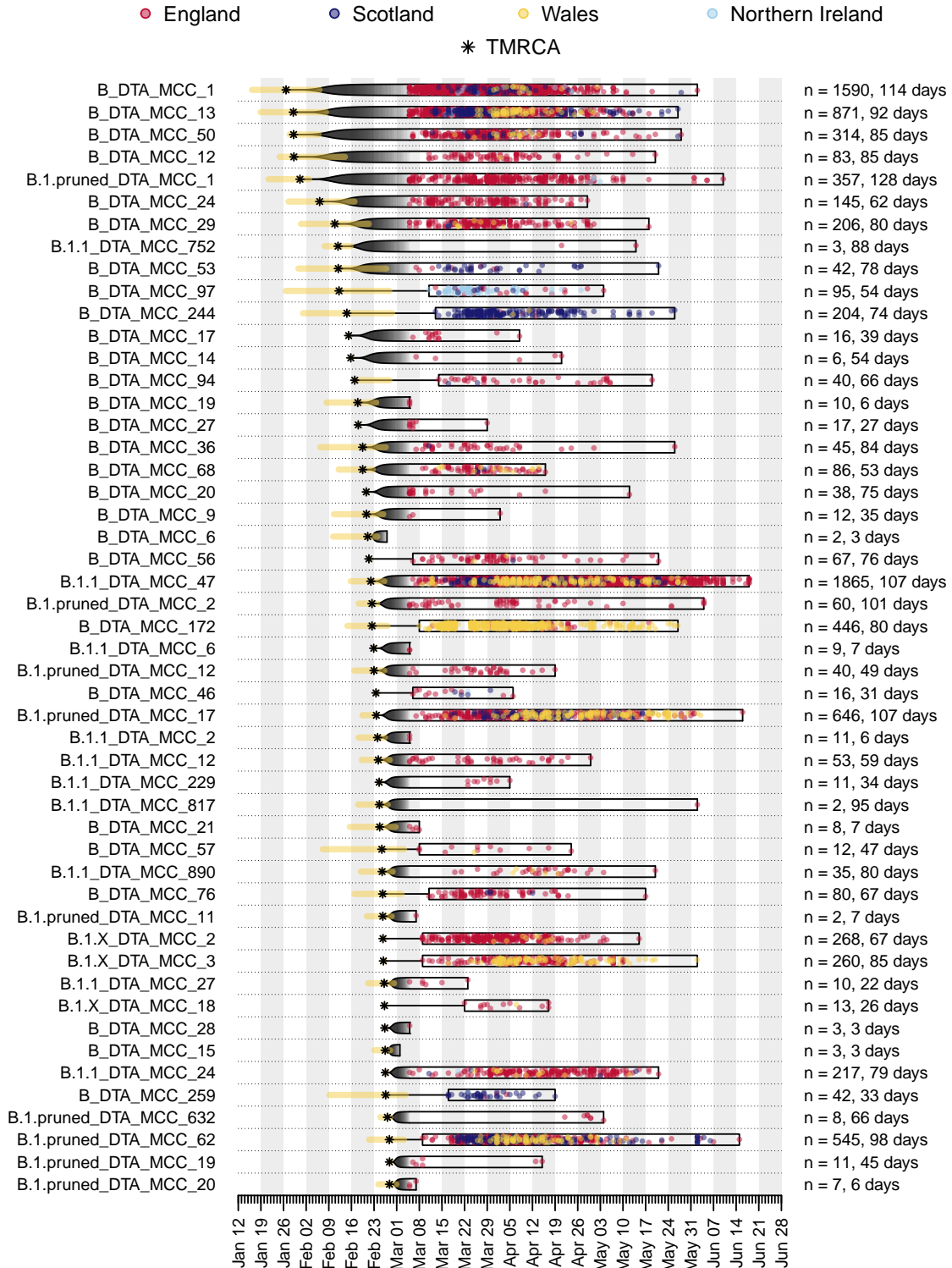


Figure 20: Illustration of the time course of the 50 earliest UK transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

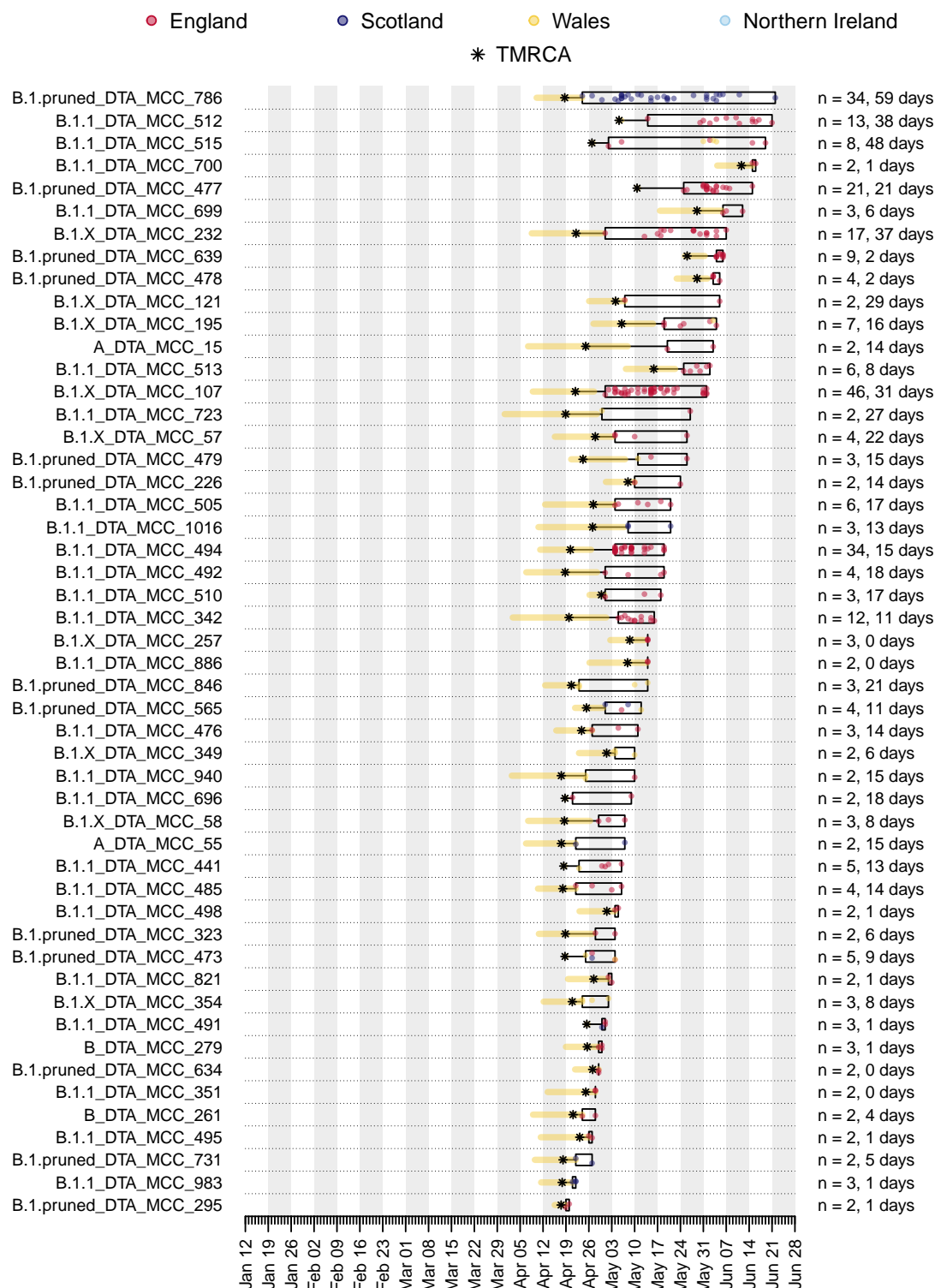


Figure 21: Illustration of the time course of the 50 most recent (by TMRCA) UK transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

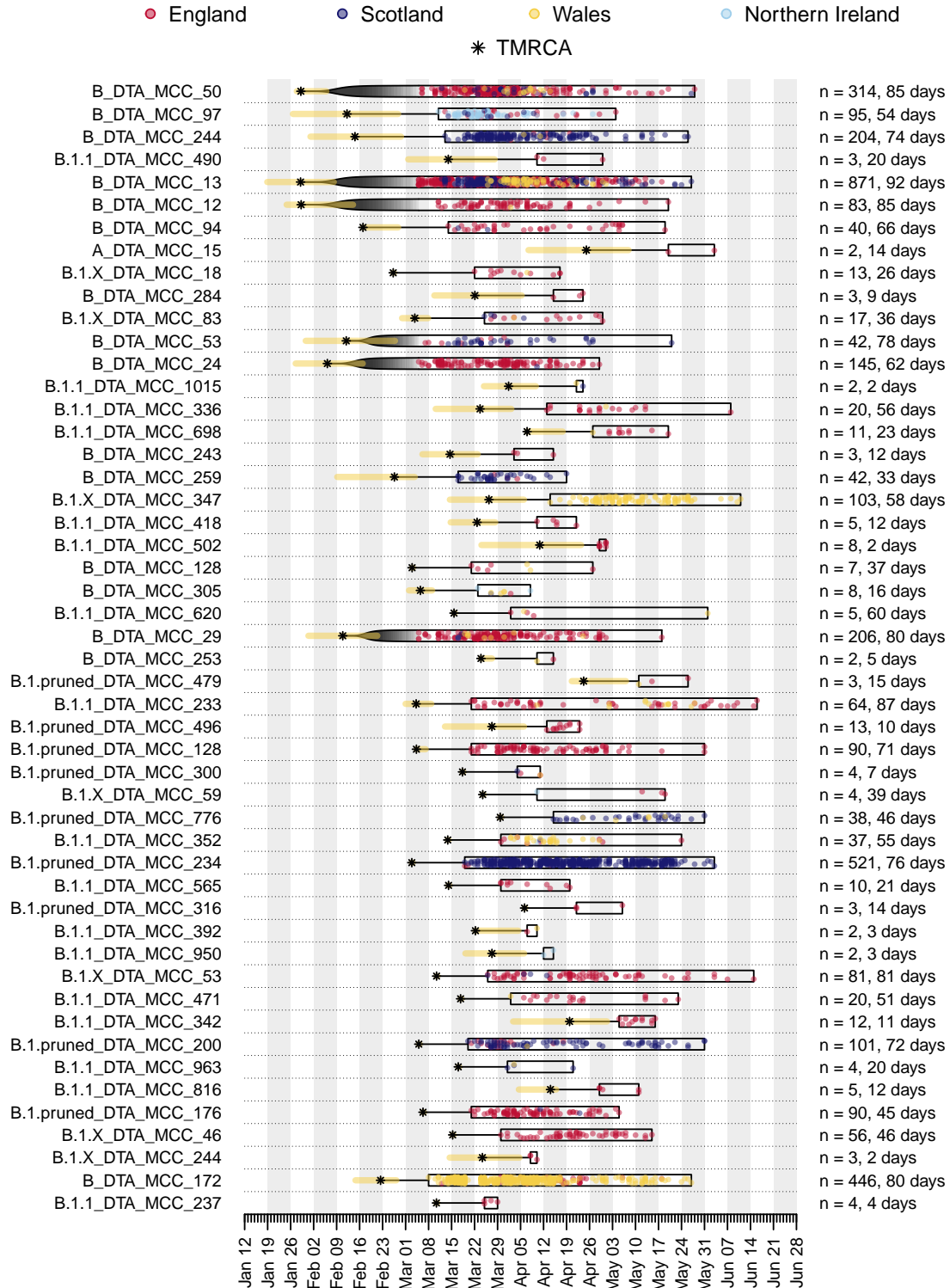


Figure 22: Illustration of the time course of the 50 UK transmission lineages in our dataset with the longest period of cryptic circulation. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

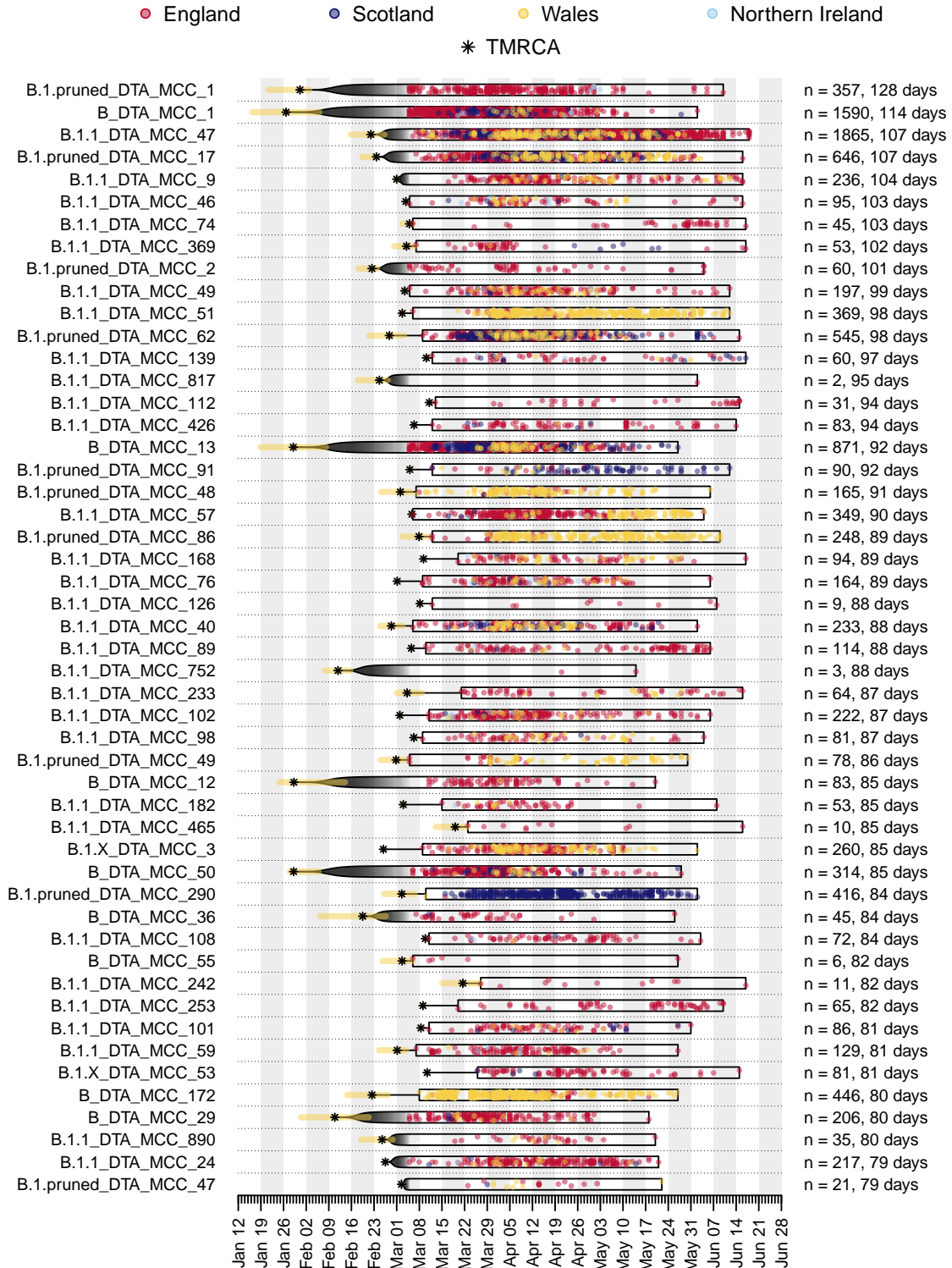


Figure 23: Illustration of the time course of the 50 UK transmission lineages in our dataset with the longest sampling duration (from earliest to most recently collected genome). Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

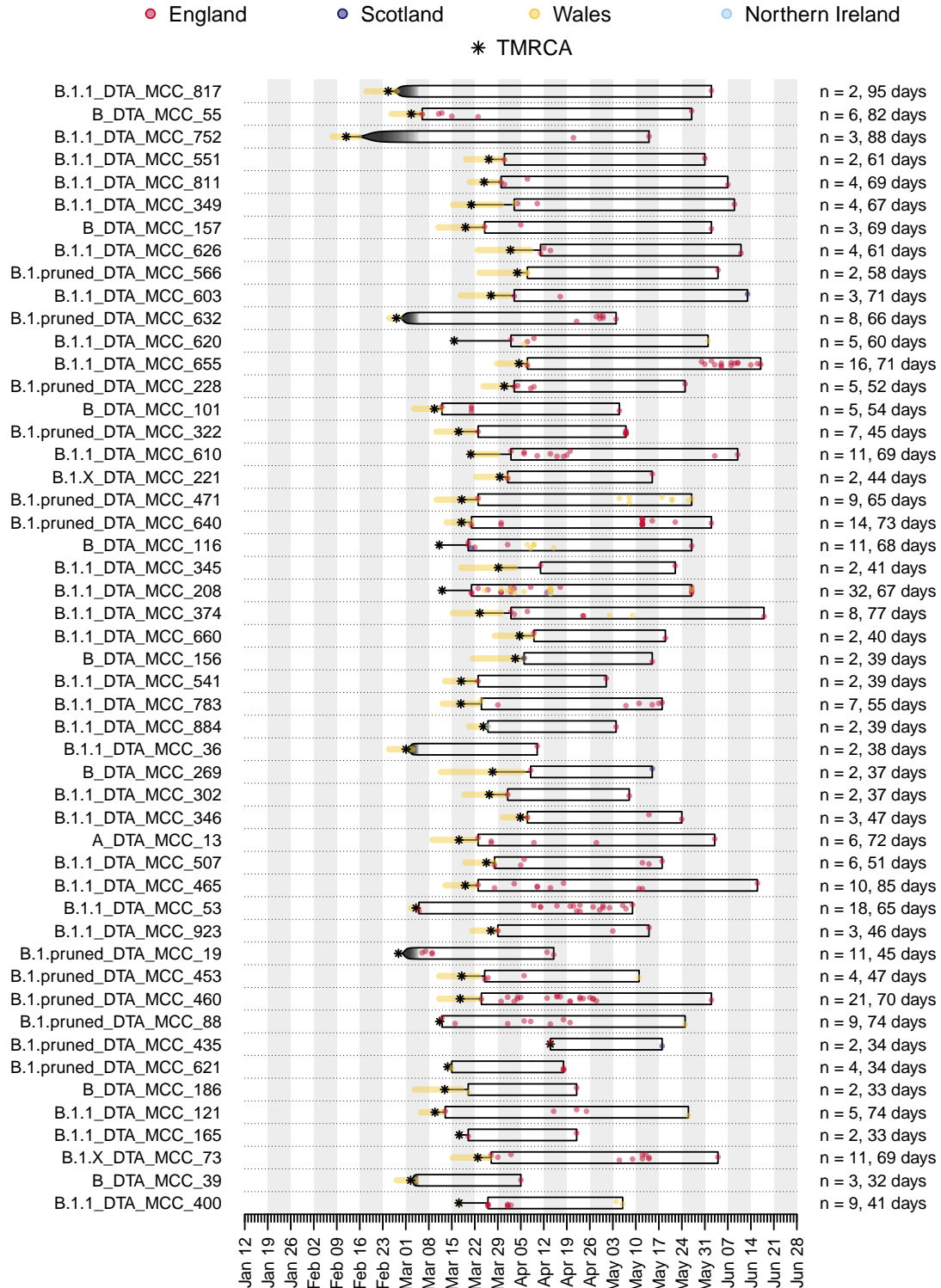


Figure 24: Illustration of the time course of the 50 UK transmission lineages in our dataset with the longest unobserved period before reemerging. Each row is a transmission lineage. Dots are genome sampling times (coloured by sampling location) and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the median TMRCA of each lineage and the yellow bars show the 95% HPD of each TMRCA. On the right, n indicates the number of UK genomes in the lineage and the duration of lineage detection (time between the lineage's oldest and most recent genomes). Sampling times of the first 500 SARS-CoV-2 genomes collected in the UK have been obscured.

6 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] tictoc_1.0          viridis_0.5.1      viridisLite_0.3.0  gplots_3.0.1.1
## [5] lubridate_1.7.4
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.3          highr_0.8          RColorBrewer_1.1-2 compiler_3.5.1
## [5] pillar_1.4.2        bitops_1.0-6       tools_3.5.1        digest_0.6.23
## [9] evaluate_0.14       lifecycle_0.1.0    tibble_2.1.3       gtable_0.3.0
## [13] pkgconfig_2.0.3     rlang_0.4.2        yaml_2.2.0         xfun_0.15
## [17] gridExtra_2.3       stringr_1.4.0      dplyr_0.8.3        knitr_1.29
## [21] gtools_3.8.1        caTools_1.17.1.3   grid_3.5.1         tidyselect_0.2.5
## [25] glue_1.3.1          R6_2.4.1           rmarkdown_2.3      gdata_2.18.0
## [29] ggplot2_3.2.1       purrr_0.3.3        magrittr_1.5       scales_1.1.0
## [33] htmltools_0.4.0     assertthat_0.2.1   colorspace_1.4-1   KernSmooth_2.23-16
## [37] stringi_1.4.3       lazyeval_0.2.2     munsell_0.5.0      crayon_1.3.4
```