

SARS-CoV-19 State introductions

Transmission Lineage Breakdown for Hamburg

Last modified: 27 Jun 2021

Contents

1	Summary	1
1.1	Data and Method	1
2	Sample breakdown (daily)	2
3	Sample breakdown into lineages (weekly)	4
4	Individual transmission lineage plots	13
4.1	Biggest transmission lineages	14
4.2	Earliest transmission lineages	16
4.3	Newest transmission lineages	18
4.4	Longest unobserved period before reactivating	22
5	Session info	24

1 Summary

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

1.1 Data and Method

- GISAID tree until 2021-06-21 as initial tree.
- The tree contains 541,000 ??? sequences.
- The small branches are collapsed.
- The tree is time-calibrated by TreeTime.
- Sankoff algorithm is used to assign location (Germany and non-Germany) to inner vertices of the tree.

2 Sample breakdown (daily)

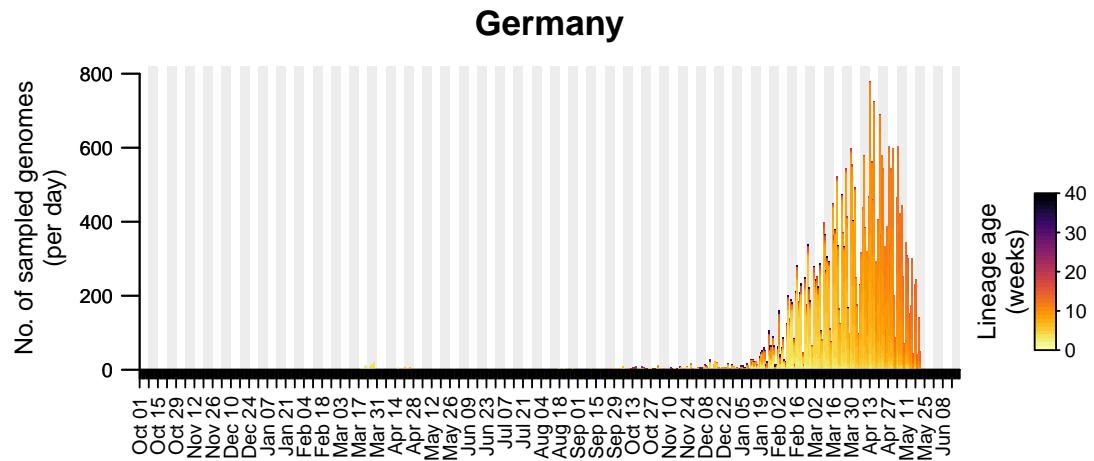


Figure 1: Number of genomes collected each day, coloured by the time since the age of the transmission lineage when the genome was collected (time from the oldest sampled genome in the lineage to the sampling time of the genome). Note that **only** genomes in transmission lineages are shown (no singletons).

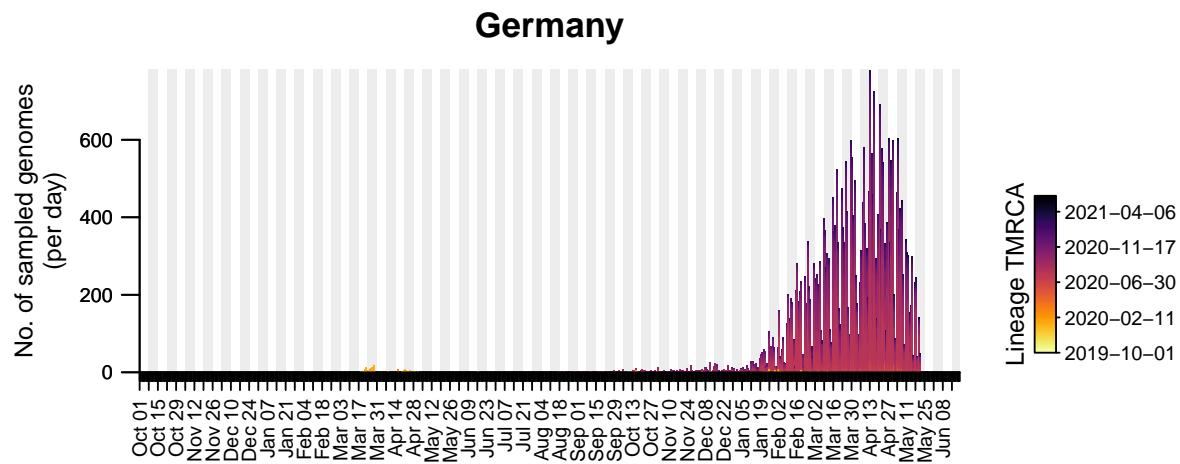


Figure 2: Number of genomes collected each day, coloured by the TMRCA of the transmission lineage. Note that **only** genomes in transmission lineages are shown (no singletons).

3 Sample breakdown into lineages (weekly)

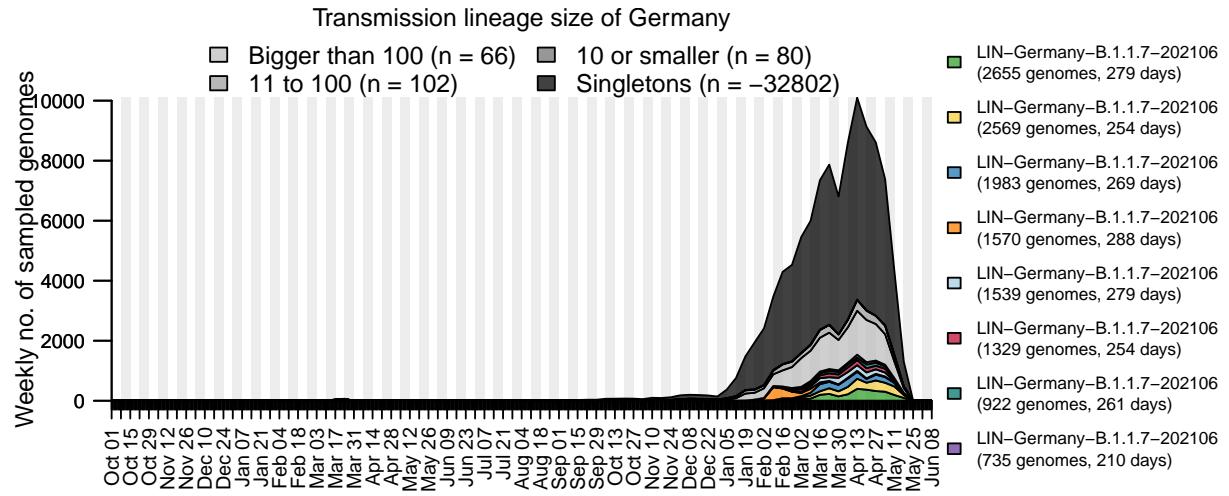
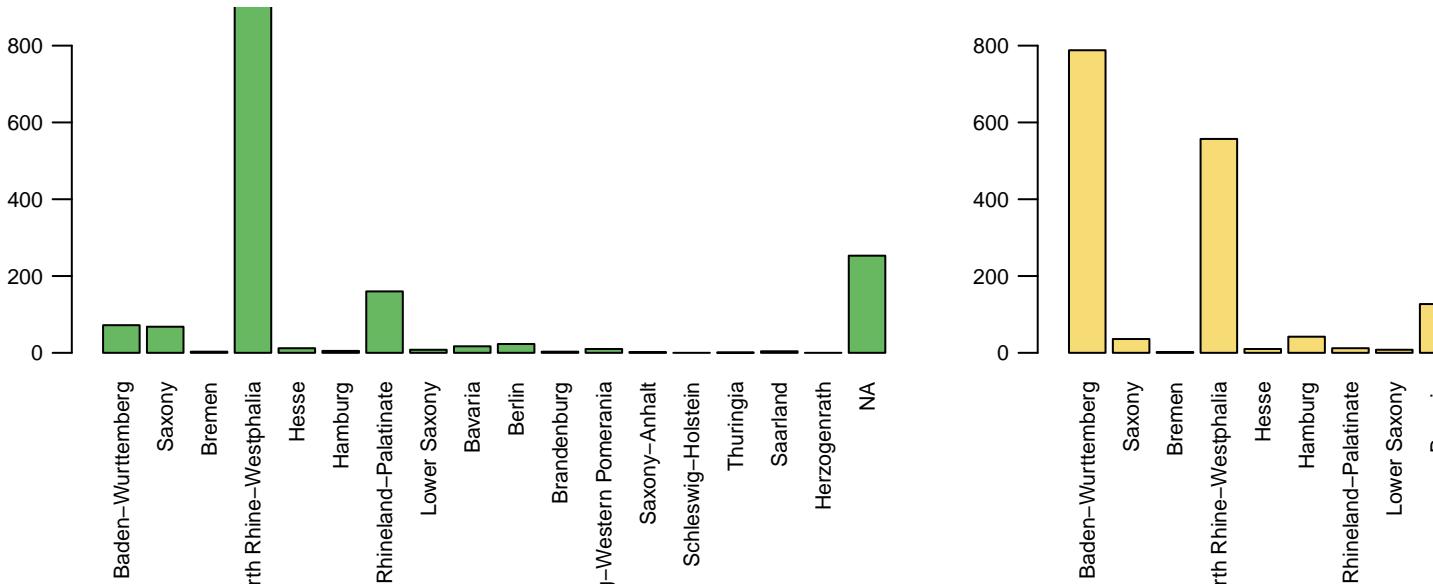


Figure 3: Lineage size breakdown of Germany genomes collected each week. The 8 largest lineages are coloured.



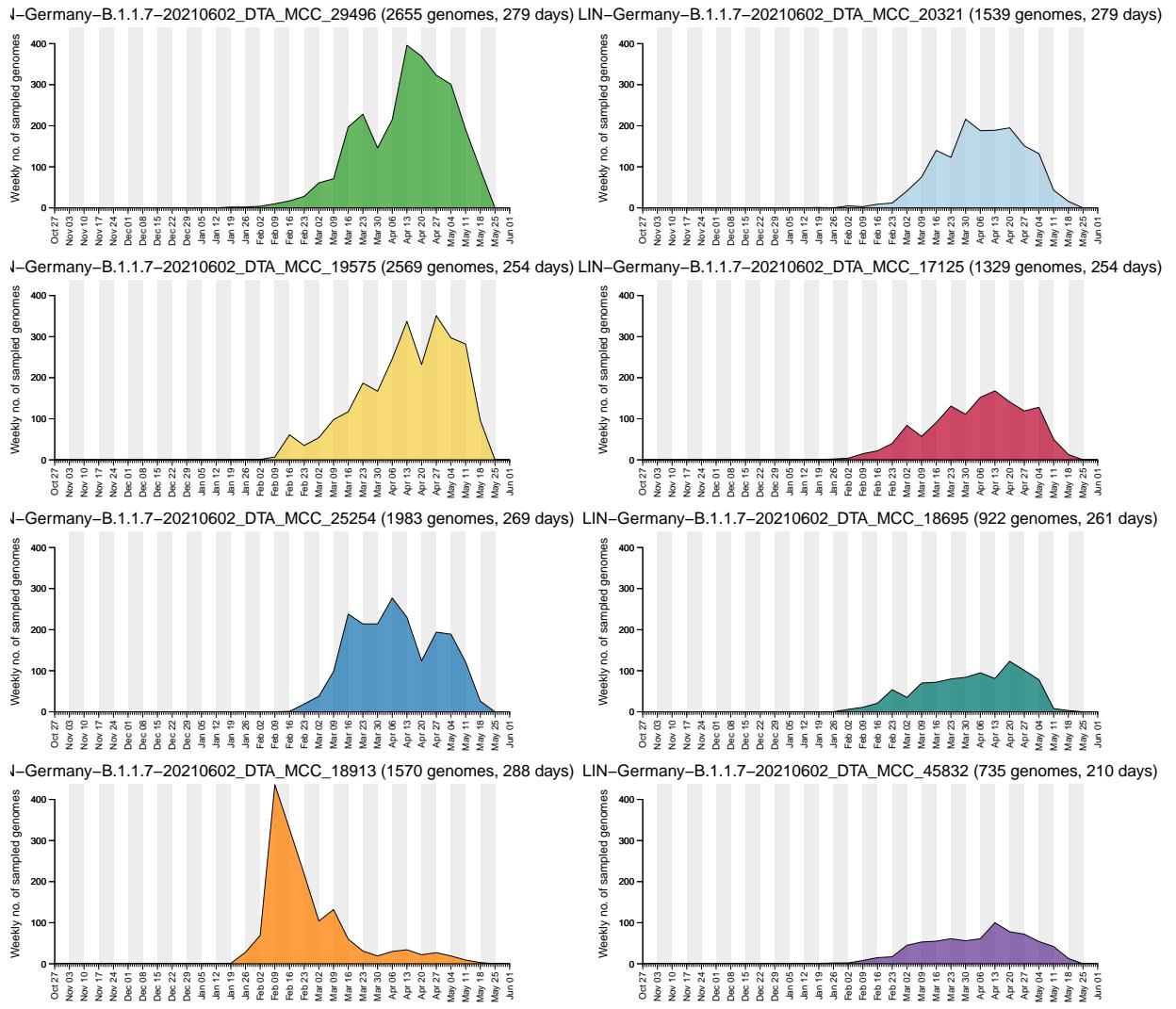


Figure 4: The weekly sampling frequency of the 8 largest `r state` transmission lineages.

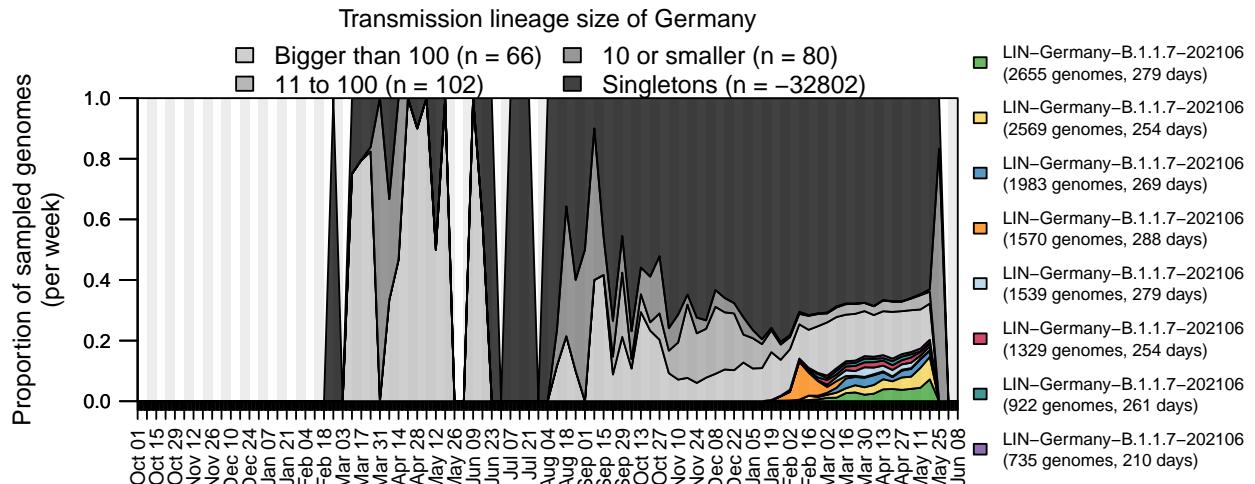
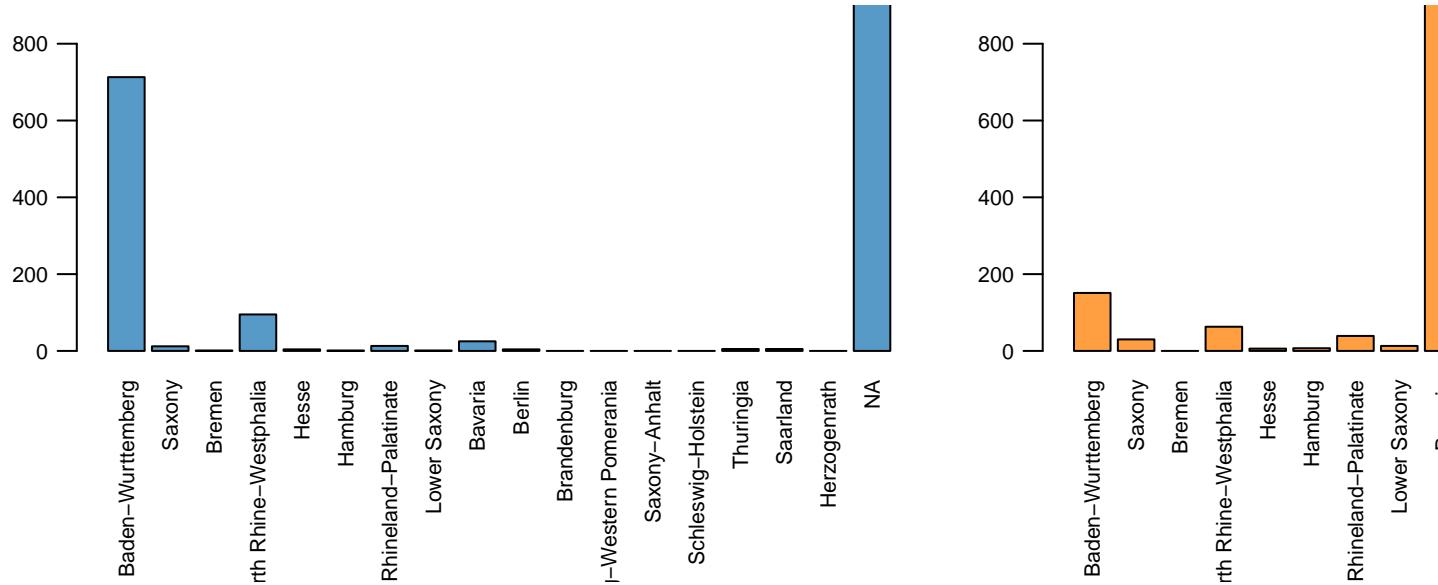
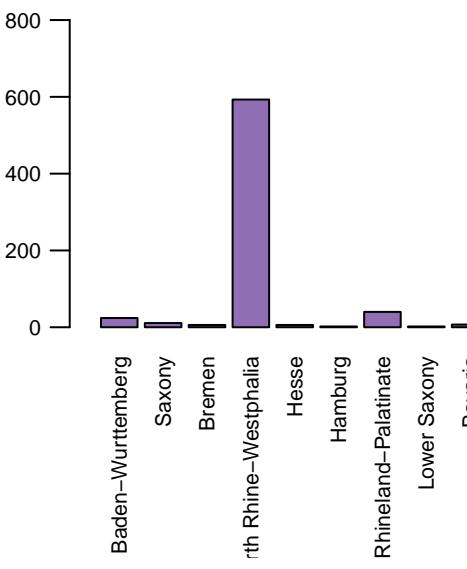
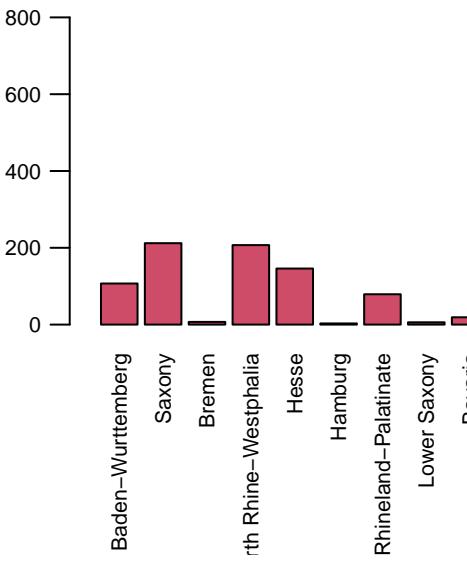
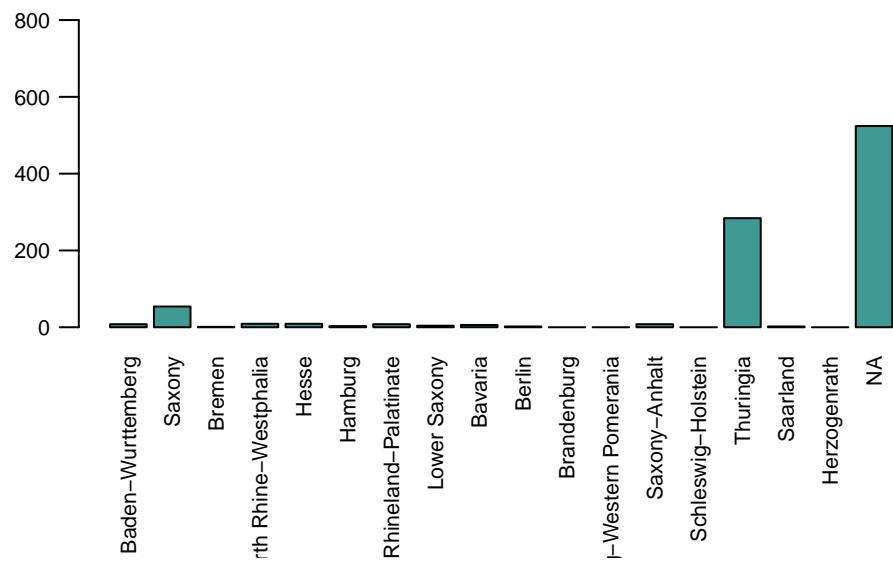
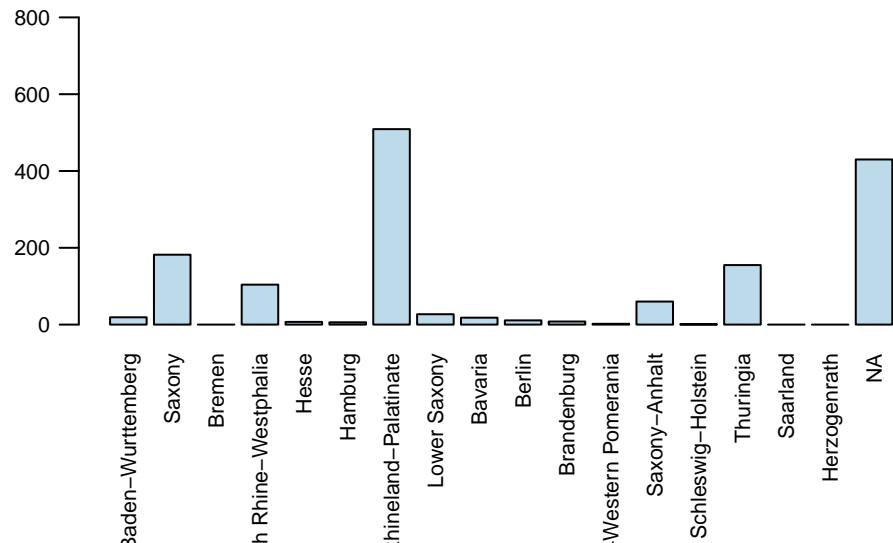
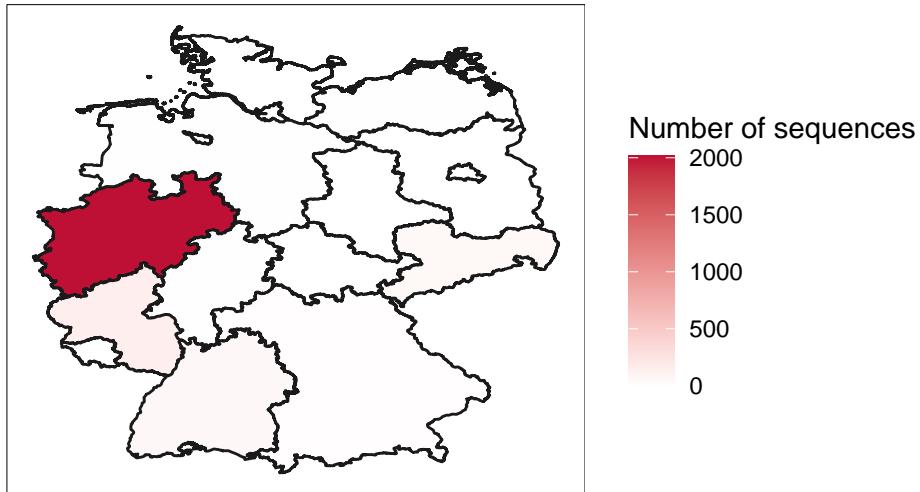


Figure 5: Lineage size breakdown of Germany genomes collected each week. The 8 largest lineages are coloured.

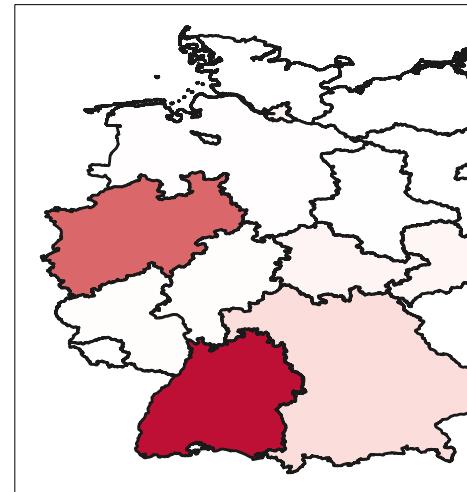




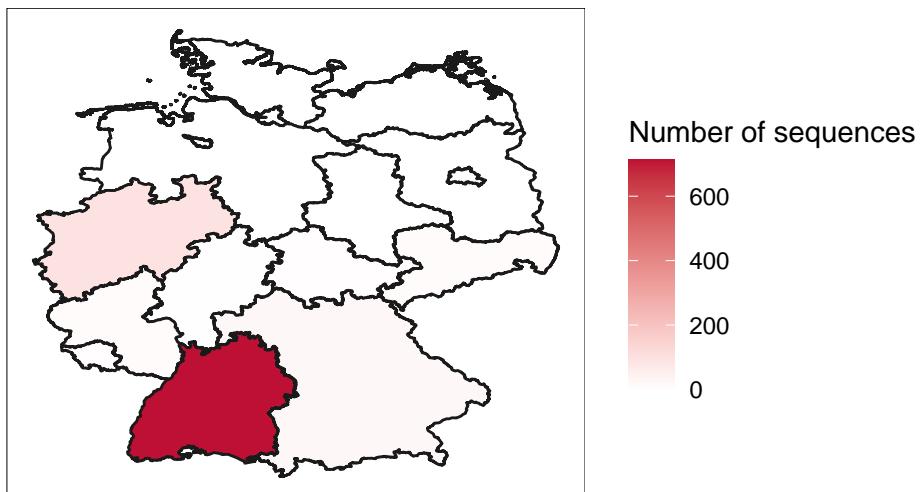
LIN-Germany-B.1.1.7-20210602_DTA_MCC_29496



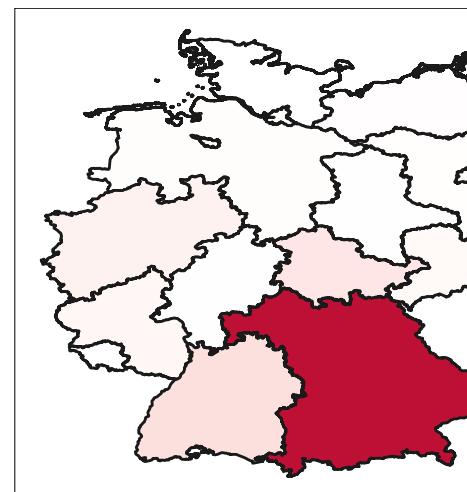
LIN-Germany-B.1.1.7-20210602_DTA_MCC_29496



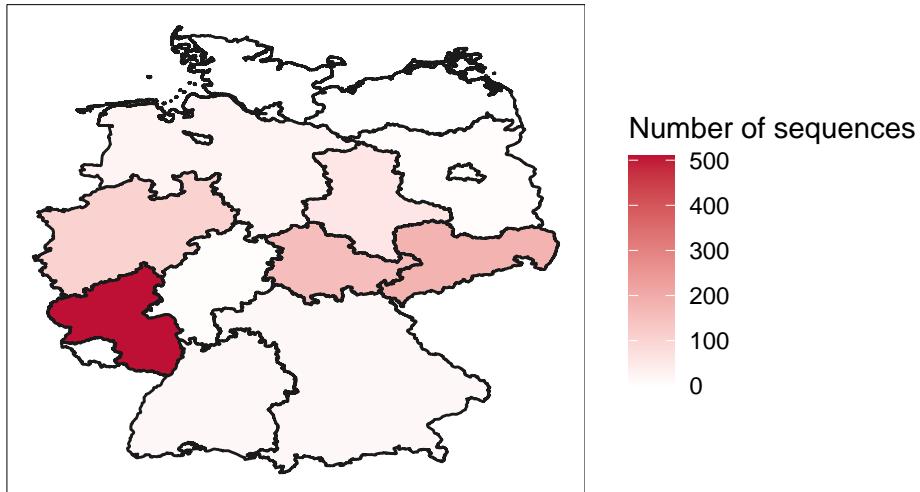
LIN-Germany-B.1.1.7-20210602_DTA_MCC_25254



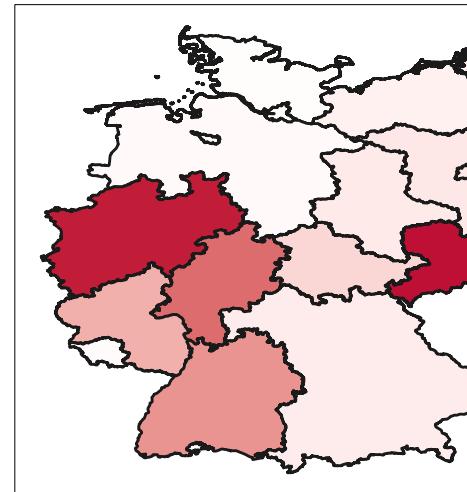
LIN-Germany-B.1.1.7-20210602_DTA_MCC_25254



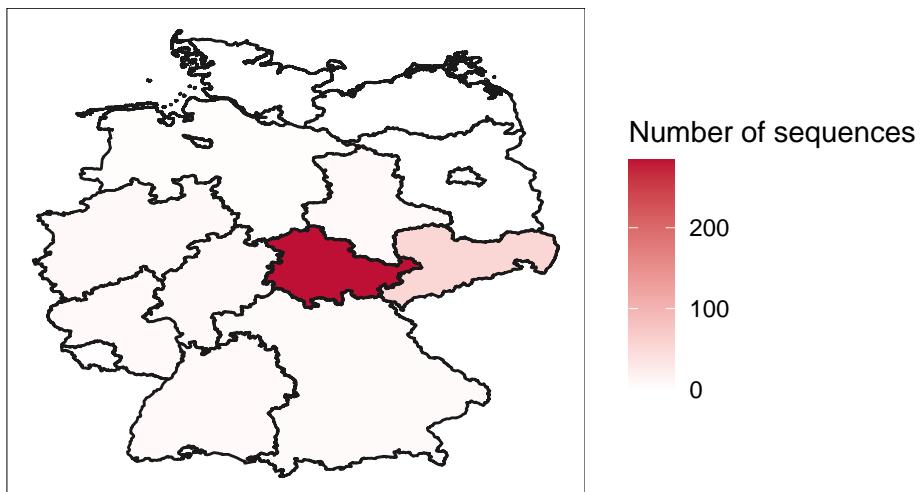
LIN-Germany-B.1.1.7-20210602_DTA_MCC_20321



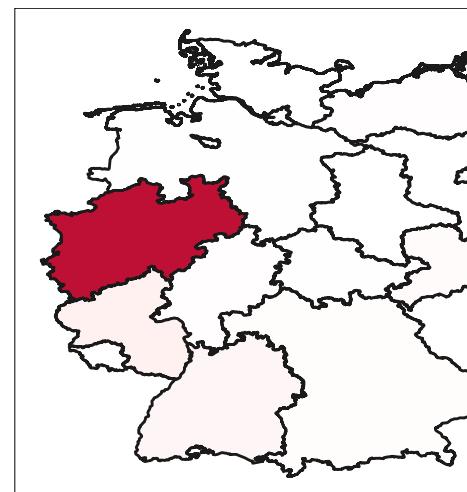
LIN-Germany-B.1.1.7-20210602_DTA_MCC_18695



LIN-Germany-B.1.1.7-20210602_DTA_MCC_18695



LIN-Germany-B.1.1.7-20210602_DTA_MCC_18695



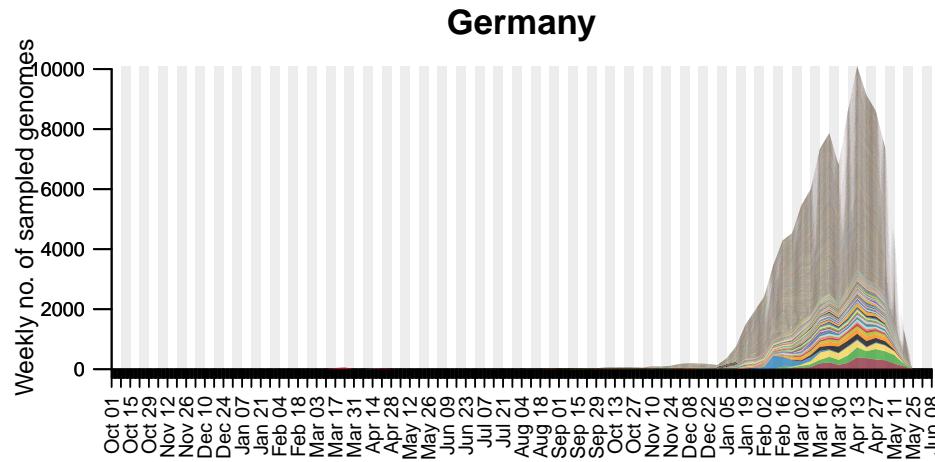


Figure 6: Lineage size breakdown of Germany genomes collected each week.

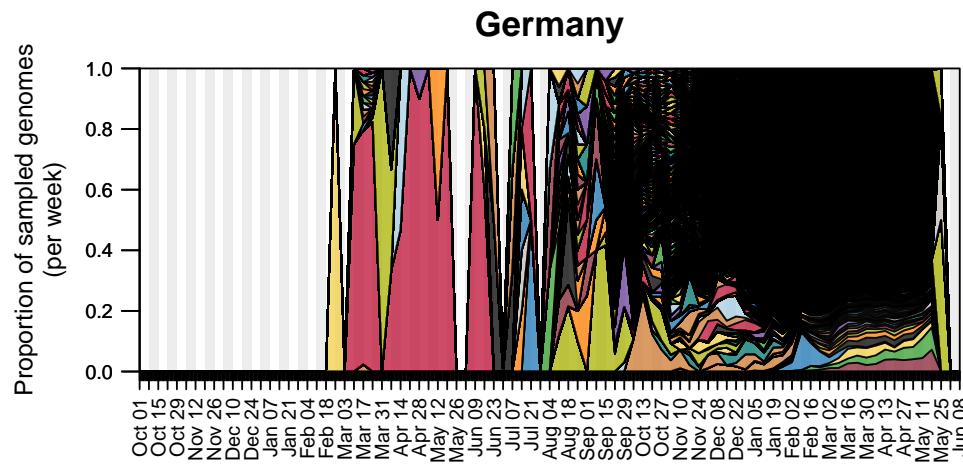


Figure 7: Lineage size breakdown of Germany genomes collected each week.

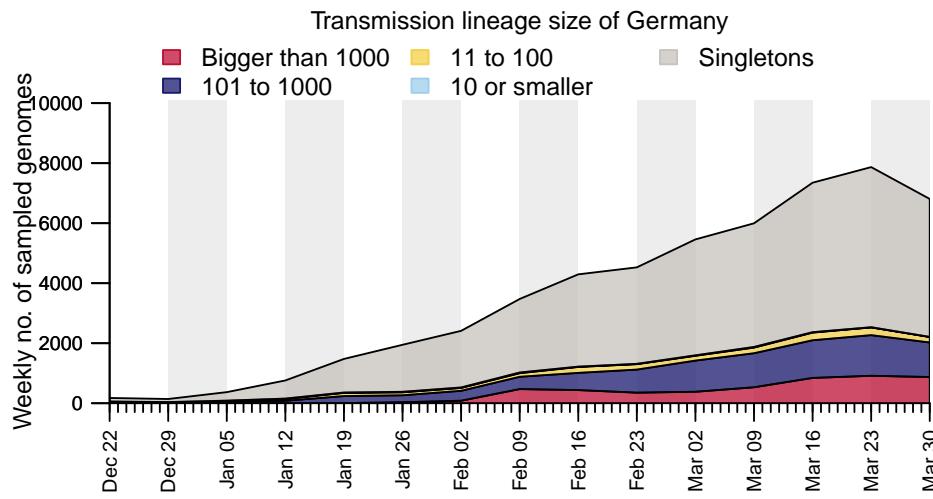
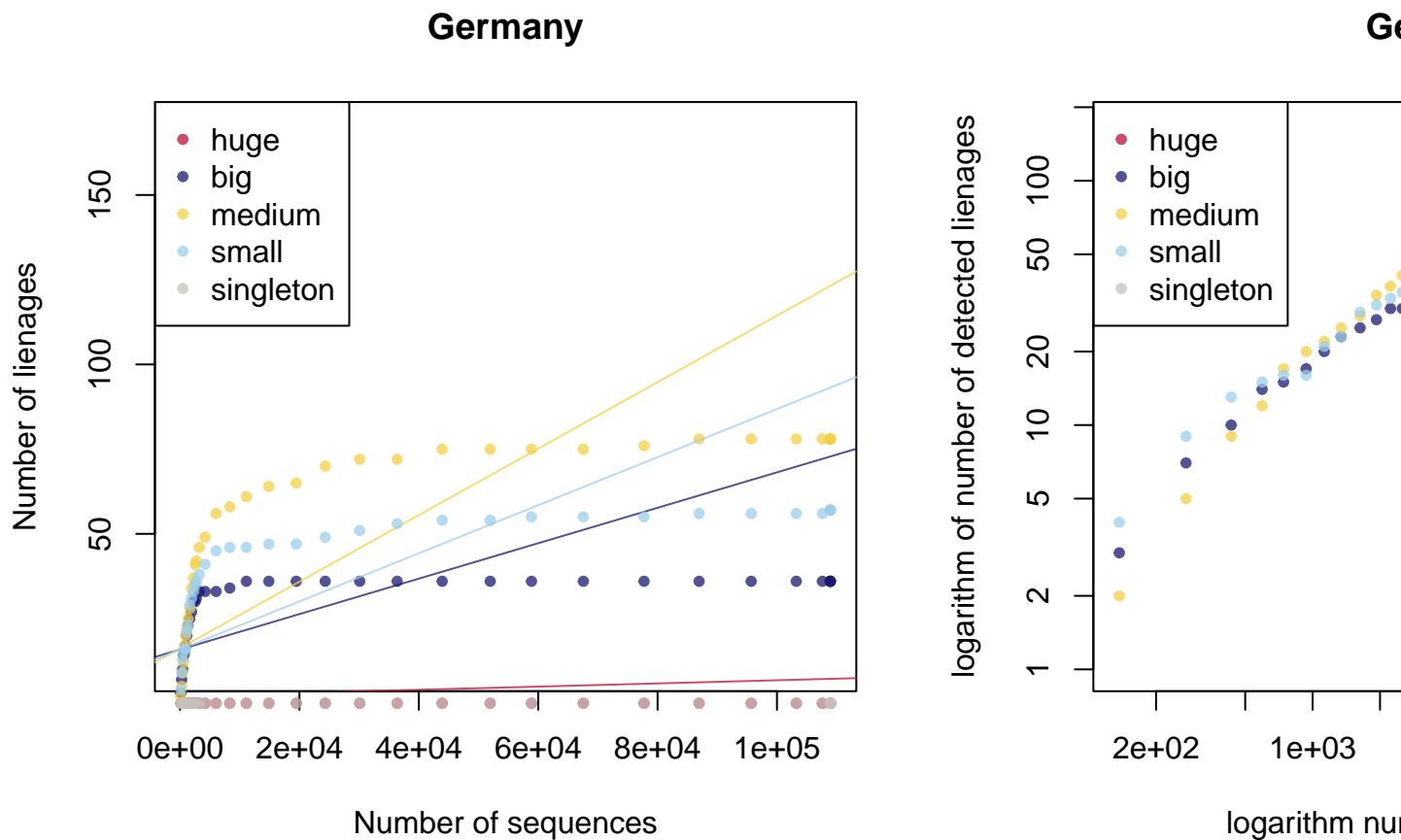


Figure 8: Lineage size breakdown of Germany genomes collected each week.

```
## Warning in xy.coords(x, y, xlabel, ylabel, log): 36 y values <= 0 omitted from
## logarithmic plot
```



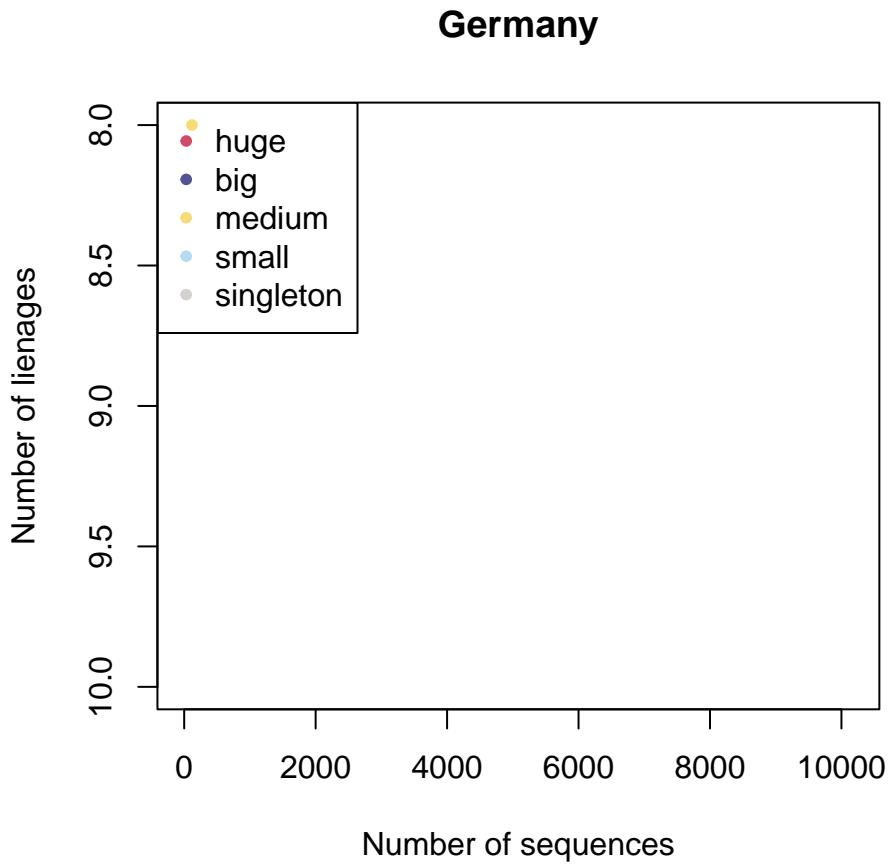


Figure 9: Lineage size breakdown of Germany genomes collected each week.

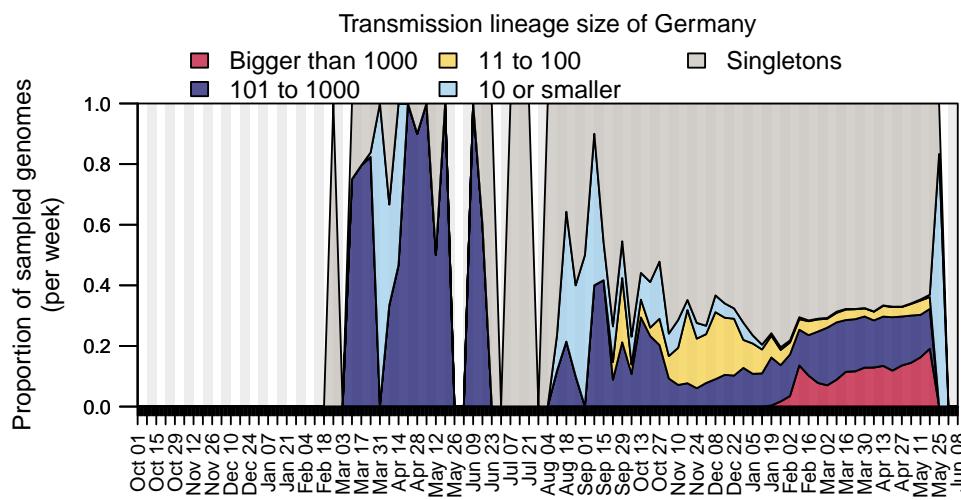


Figure 10: Lineage size breakdown of Germany genomes collected each week.

4 Individual transmission lineage plots

4.1 Biggest transmission lineages

```
## Error in kable(clusterSamplesMCC[clusterSamplesMCC$cluster %in% bigLineages$cluster, : could not find
```

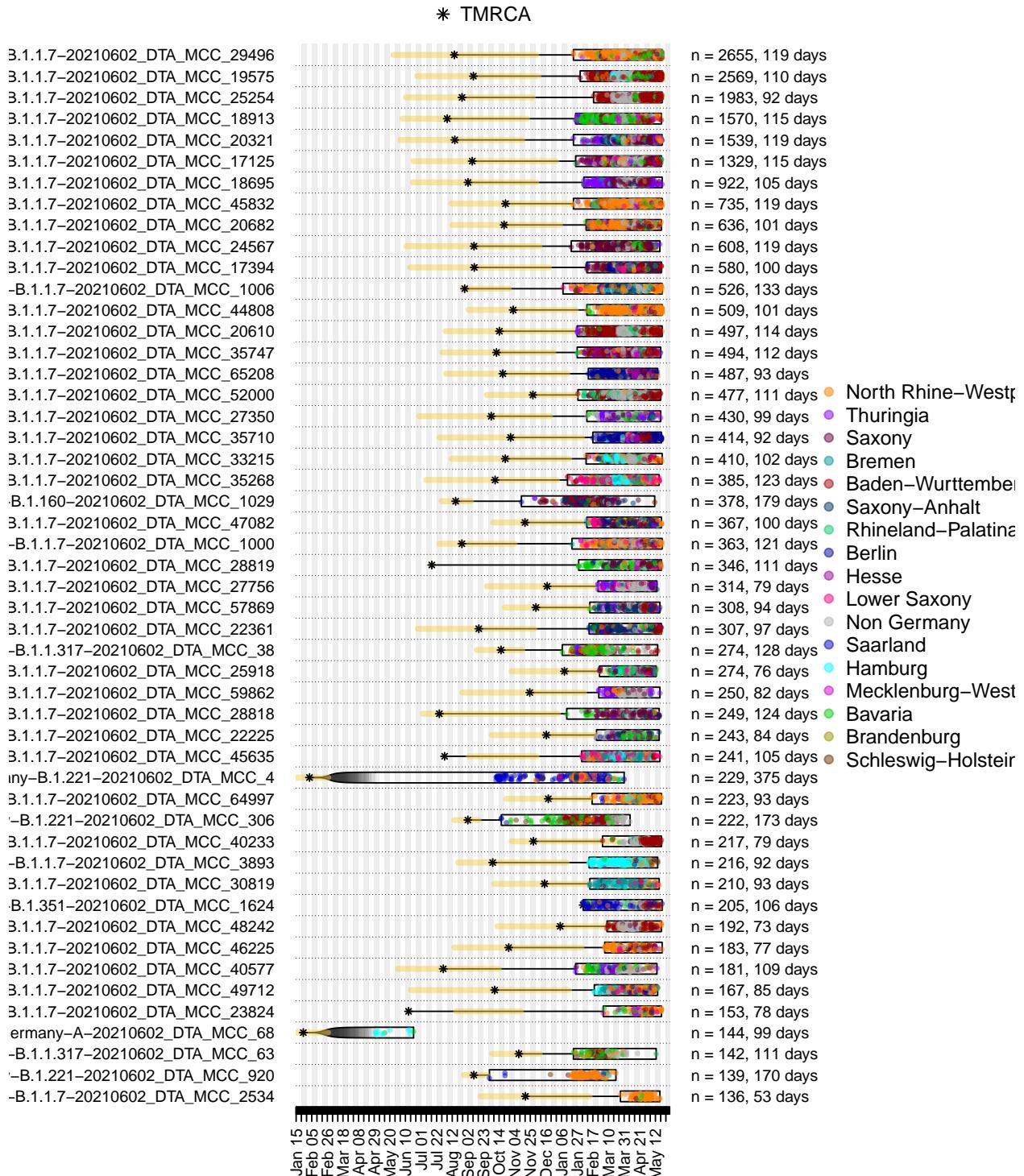


Figure 11: Illustration of the time course of the 50 largest Düsseldorf transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the TMRCA of each lineage. On the right, n indicates the number of Düsseldorf 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 Düsseldorf have been obscured.

4.2 Earliest transmission lineages

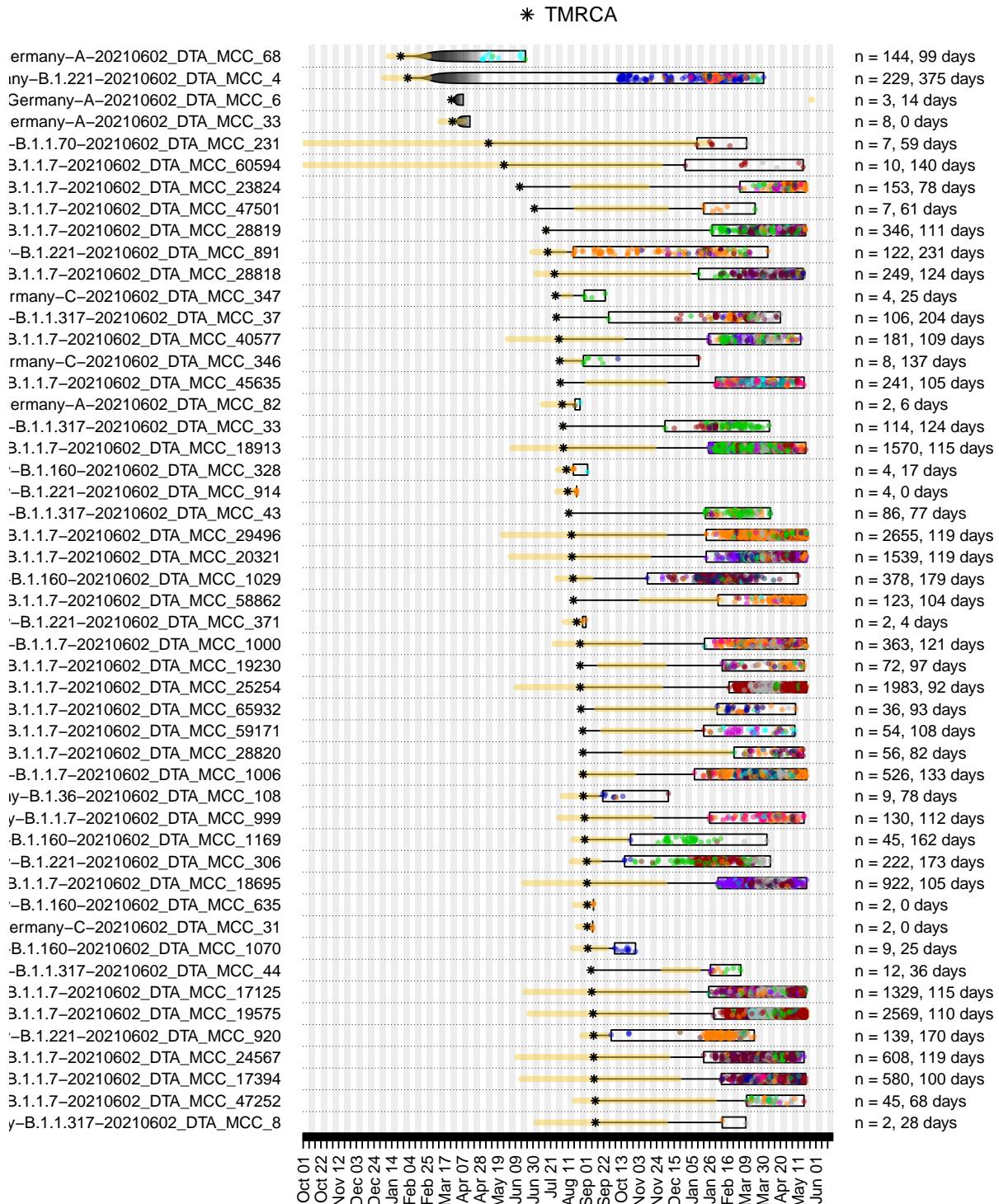
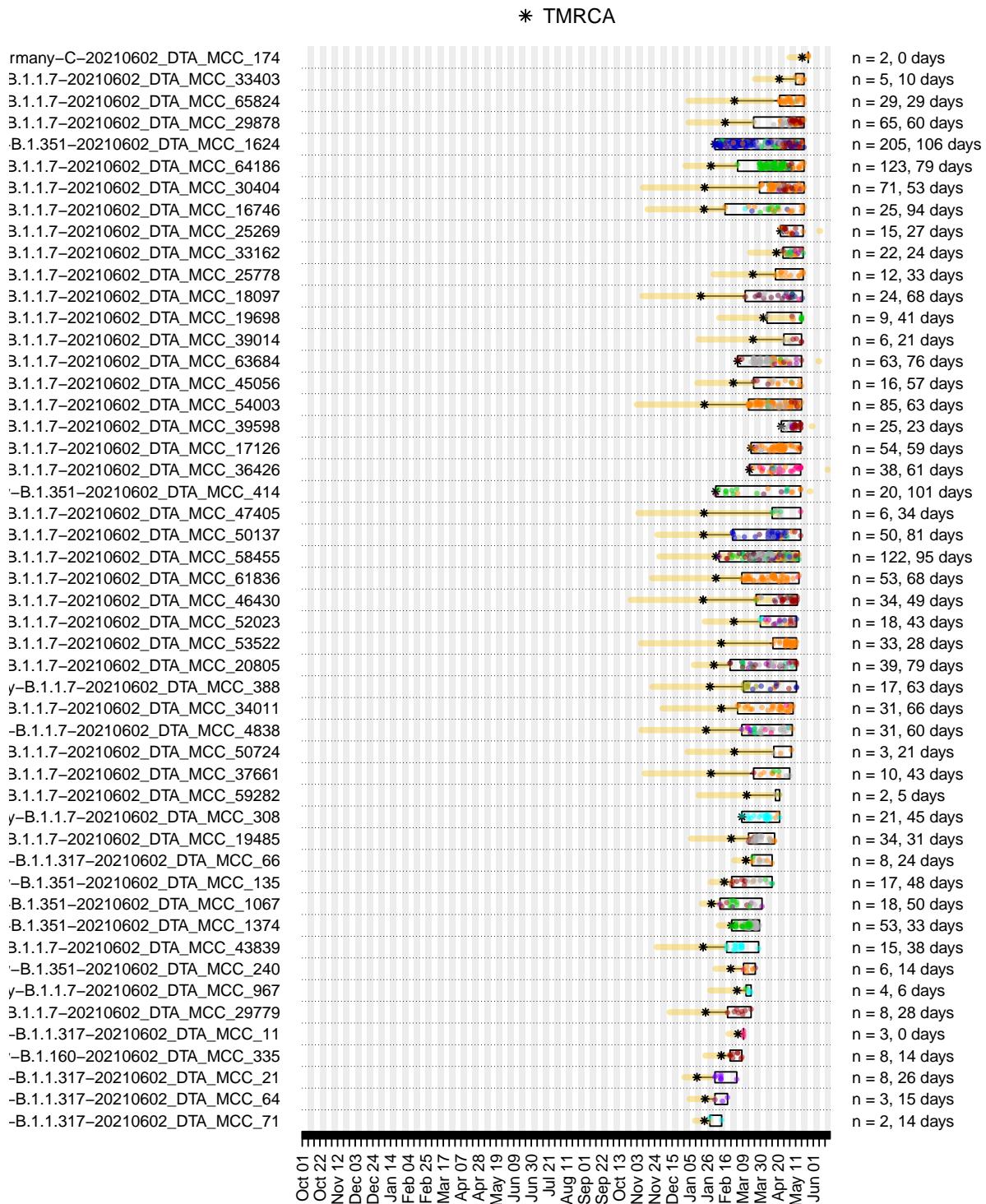
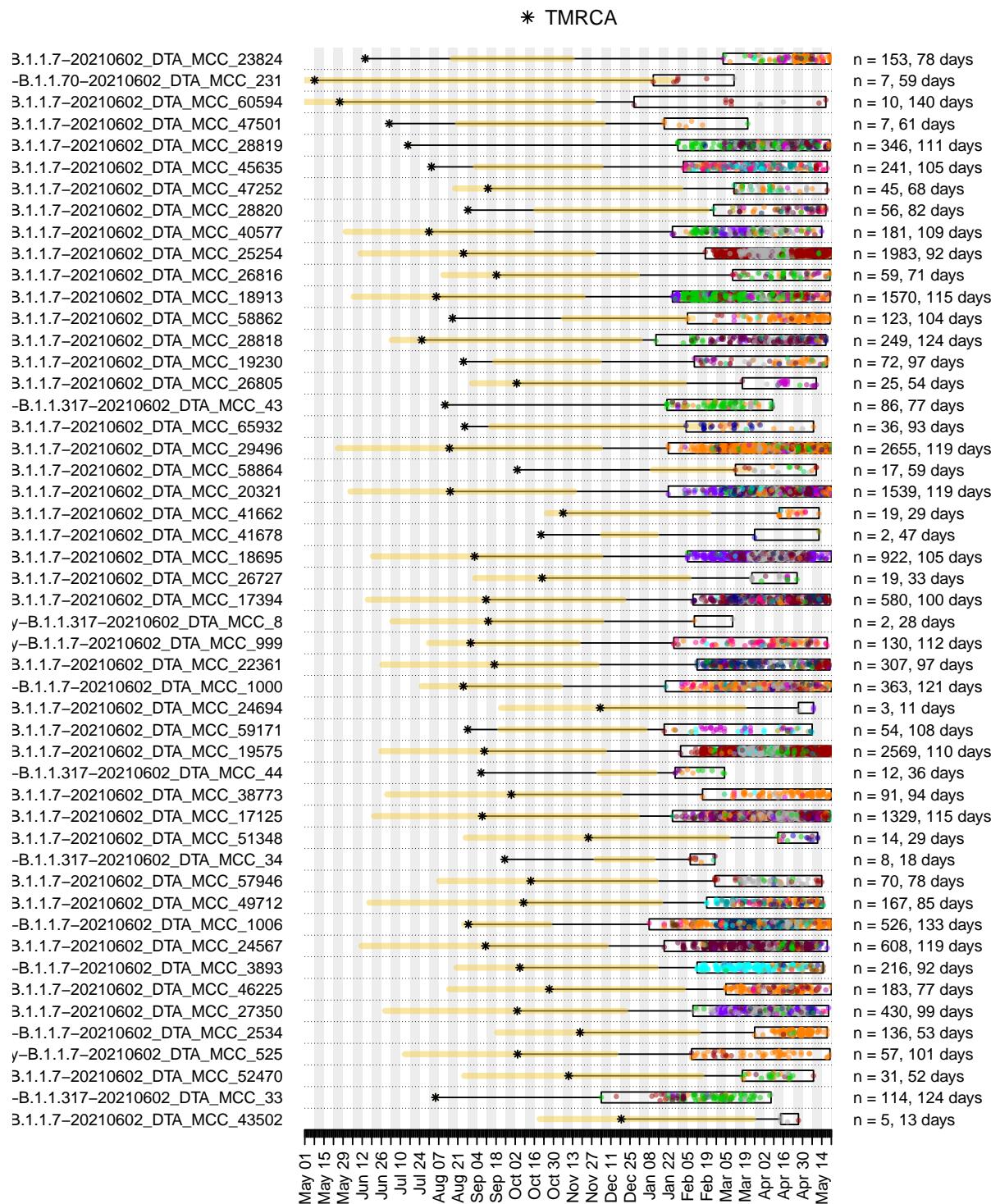


Figure 12: Illustration of the time course of the 50 earliest Dusseldorf transmission lineages in our dataset. Each row is a transmission lineage. Dots are genome sampling times and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the TMRCA of each lineage. On the right, n indicates the number of Dusseldorf 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 Dusseldorf have been obscured.

4.3 Newest transmission lineages



Longest periods of cryptic circulation



Longest sampling period

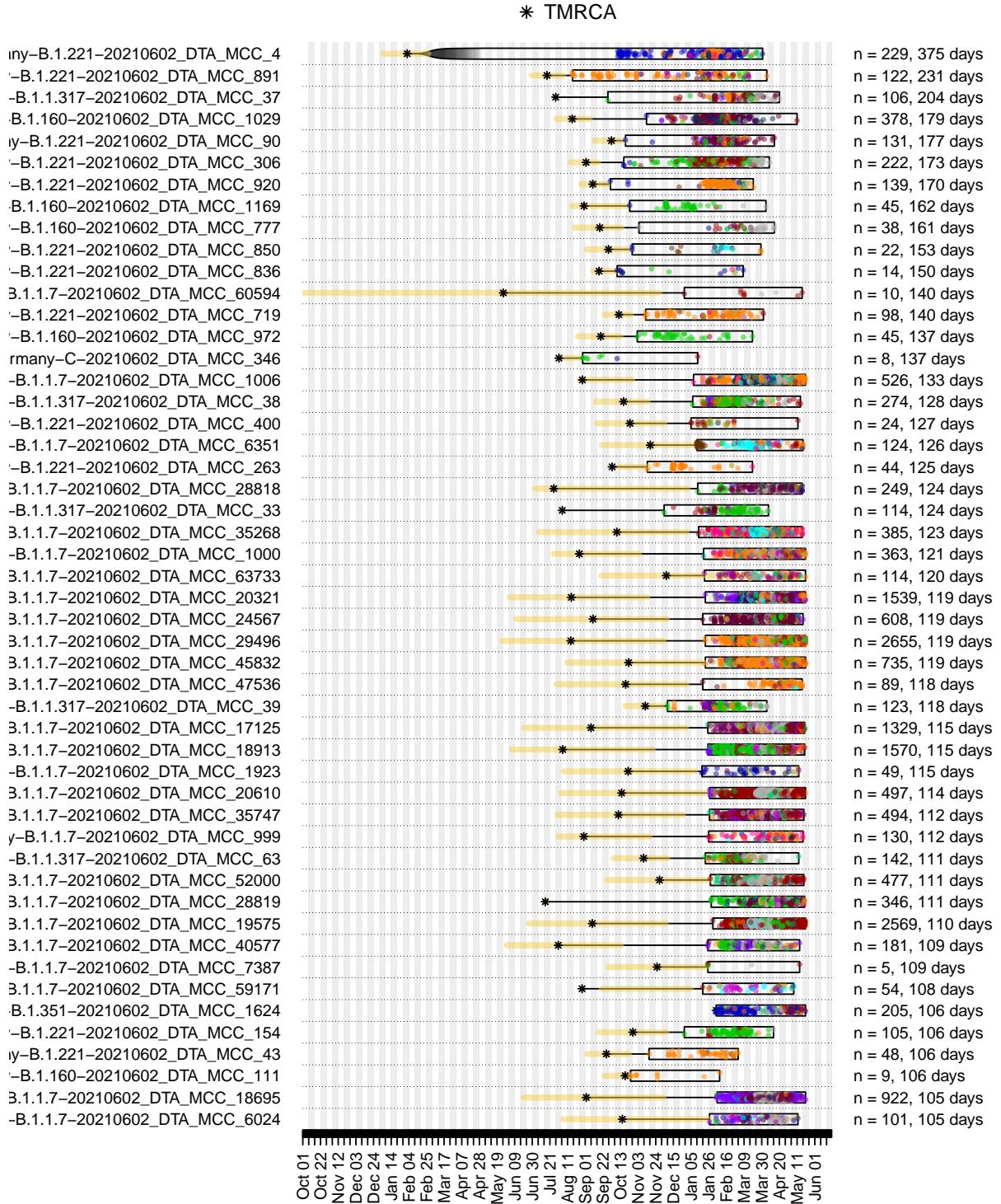


Figure 13: Illustration of the time course of the 50 Dusseldorf transmission lineages in our dataset with the longest sampling duration (from earliest to most recently collected genome). Each row is a transmission lineage. Dots are genome sampling times and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the TMRCA of each lineage. On the right, n indicates the number of Dusseldorf 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 Dusseldorf have been obscured.

4.4 Longest unobserved period before reactivating

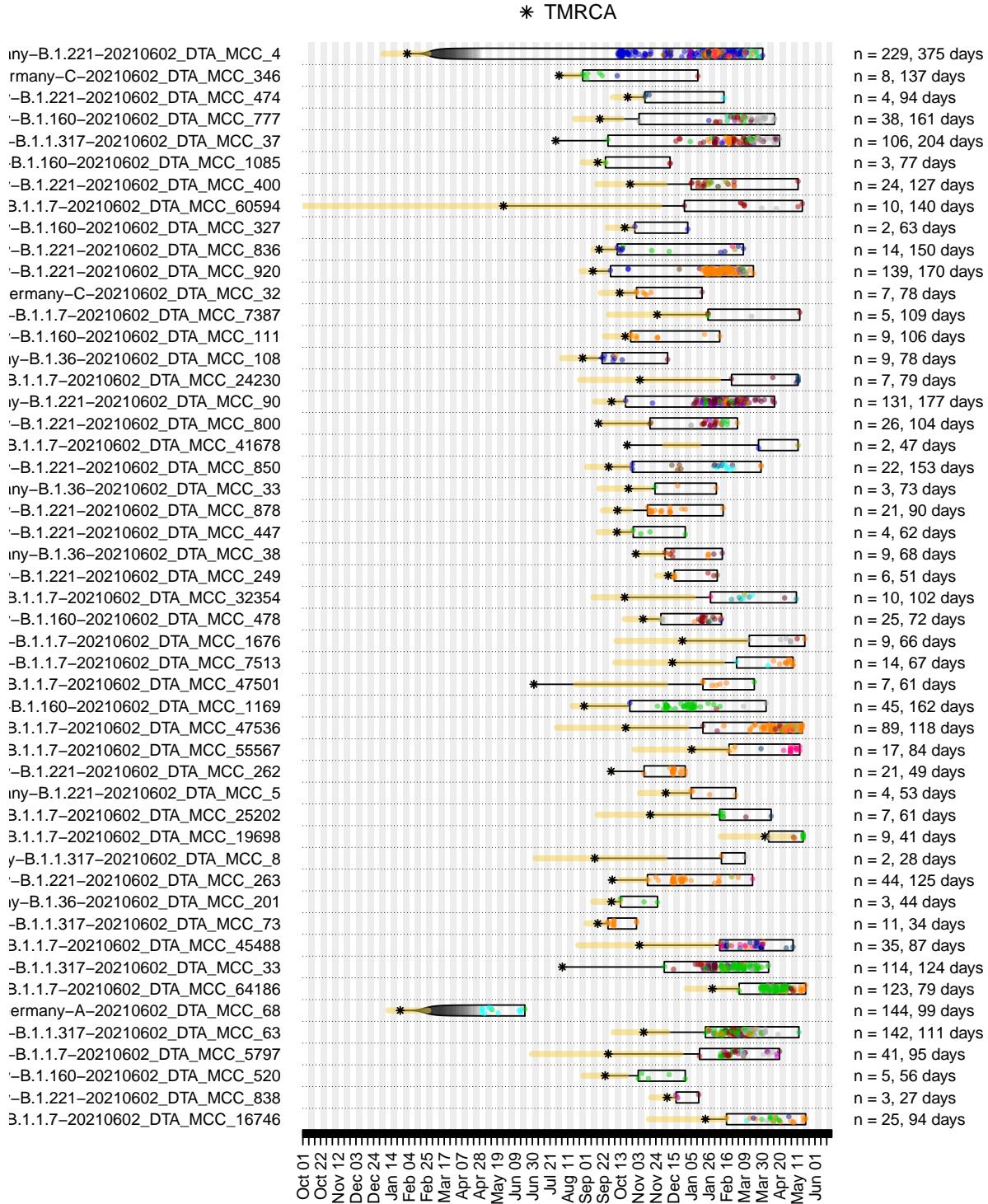


Figure 14: Illustration of the time course of the 50 Dusseldorf transmission lineages in our dataset with the longest unobserved period before reemerging. Each row is a transmission lineage. Dots are genome sampling times and boxes show the range of sampling times for each transmission lineage (sampling duration). Asterisks show the TMRCA of each lineage. On the right, n indicates the number of Dusseldorf 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 Dusseldorf have been obscured.²³

5 Session info

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 20.04.2 LTS
##
## Matrix products: default
## BLAS:    /usr/lib/x86_64-linux-gnu/atlas/libblas.so.3.10.3
## LAPACK:  /usr/lib/x86_64-linux-gnu/atlas/liblapack.so.3.10.3
##
## locale:
## [1] LC_CTYPE=en_US.UTF-8          LC_NUMERIC=C
## [3] LC_TIME=en_US.UTF-8          LC_COLLATE=en_US.UTF-8
## [5] LC_MONETARY=en_US.UTF-8       LC_MESSAGES=en_US.UTF-8
## [7] LC_PAPER=en_US.UTF-8          LC_NAME=C
## [9] LC_ADDRESS=C                  LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8   LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics   grDevices utils      datasets   methods    base
##
## other attached packages:
## [1] maptools_1.1-1    colorspace_2.0-1   ggplot2_3.3.3    raster_3.4-10
## [5] sp_1.4-5         dplyr_1.0.6      stringr_1.4.0    tictoc_1.0.1
## [9] viridis_0.6.1    viridisLite_0.4.0 gplots_3.1.1     lubridate_1.7.10
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.6        highr_0.9        RColorBrewer_1.1-2 compiler_4.1.0
## [5] pillar_1.6.1      bitops_1.0-7      tools_4.1.0       digest_0.6.27
## [9] lattice_0.20-44   evaluate_0.14    lifecycle_1.0.0   tibble_3.1.2
## [13] gtable_0.3.0     pkgconfig_2.0.3   rlang_0.4.11     yaml_2.2.1
## [17] xfun_0.23        gridExtra_2.3    withr_2.4.2      knitr_1.33
## [21] rgeos_0.5-5      generics_0.1.0   vctrs_0.3.8      gtools_3.8.2
## [25] caTools_1.18.2   tidyselect_1.1.1  grid_4.1.0       glue_1.4.2
## [29] R6_2.5.0         fansi_0.5.0      foreign_0.8-81   rmarkdown_2.8
## [33] farver_2.1.0     purrr_0.3.4      magrittr_2.0.1    codetools_0.2-18
## [37] scales_1.1.1     htmltools_0.5.1.1 ellipsis_0.3.2   labeling_0.4.2
## [41] KernSmooth_2.23-20 utf8_1.2.1     stringi_1.6.2   munsell_0.5.0
## [45] crayon_1.4.1
```