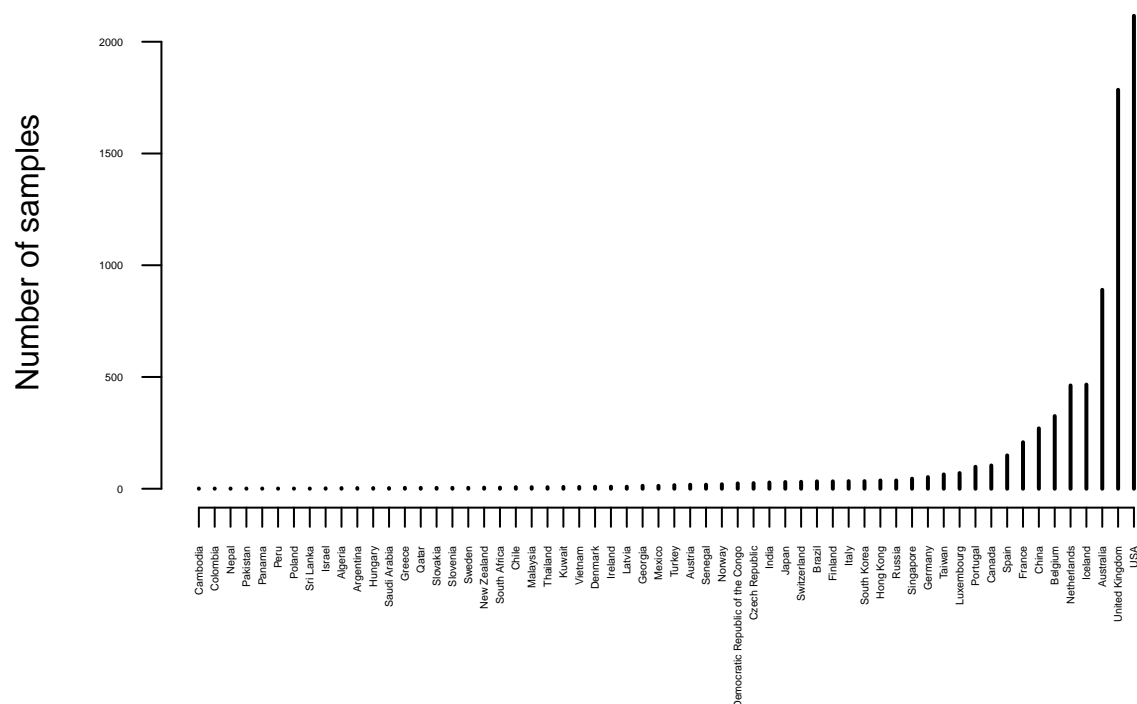


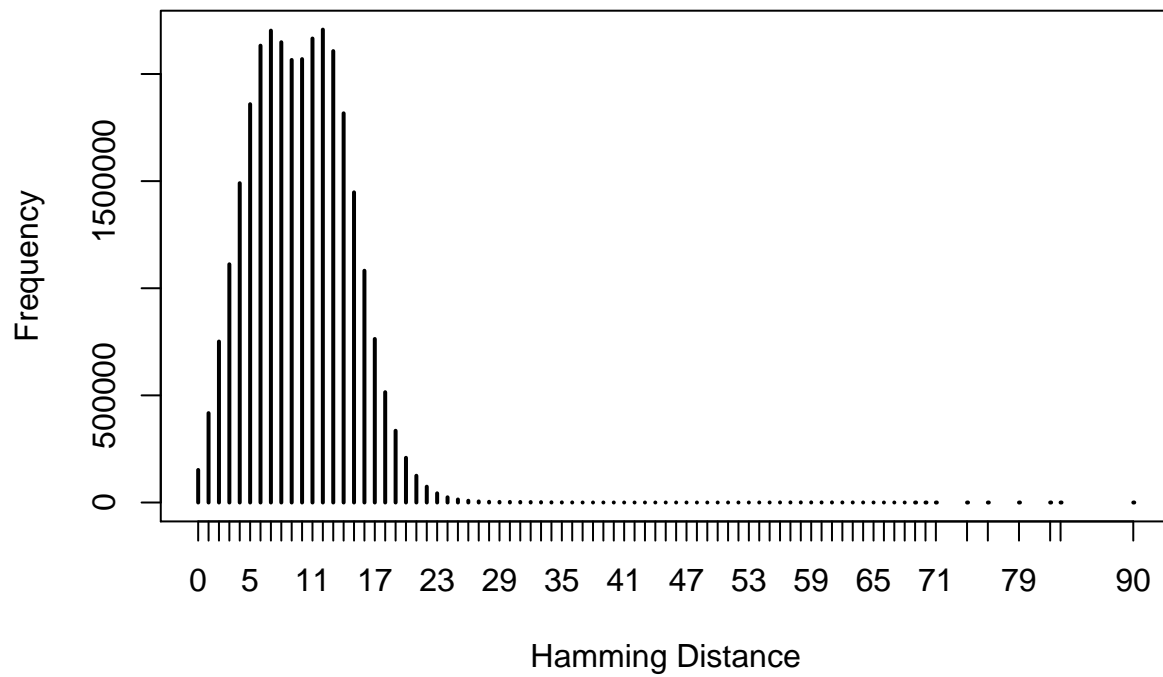
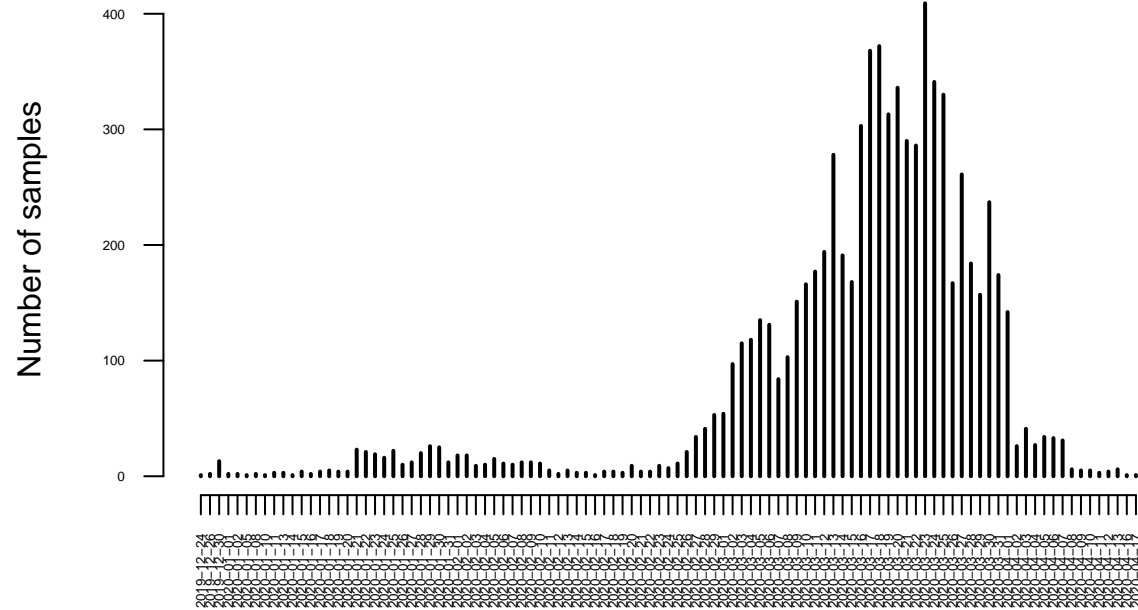
# Preliminary Molecular Clock Check and Summary Statistics

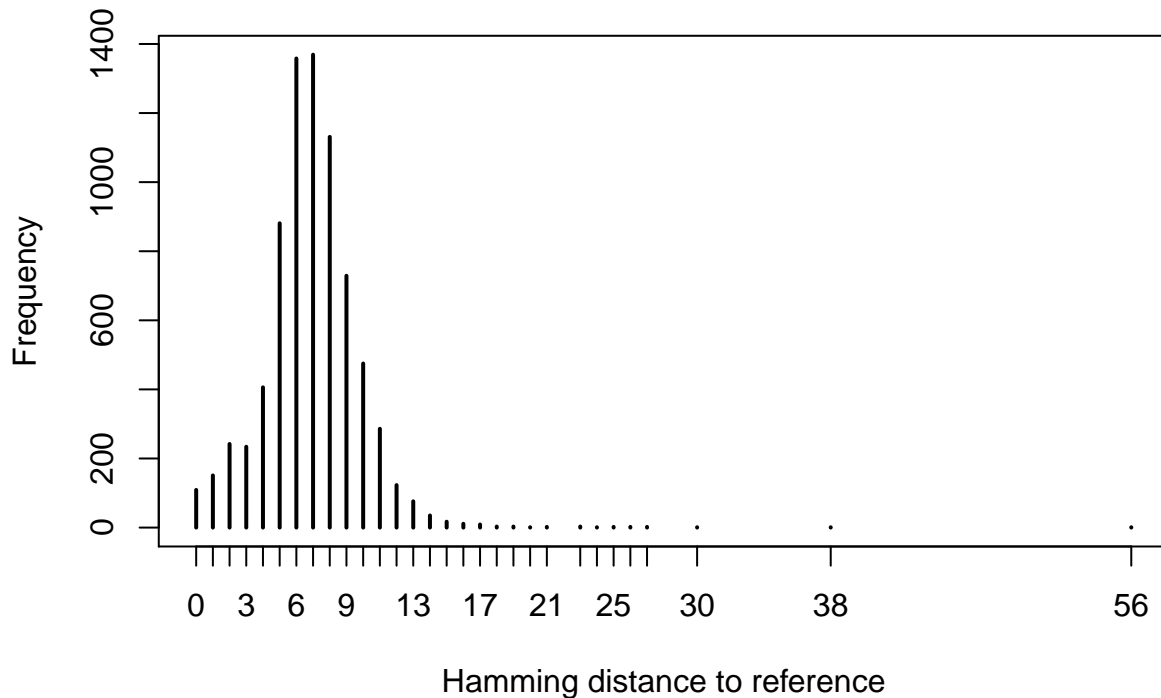
```
##
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':
##
## date
```



## Sequences by date





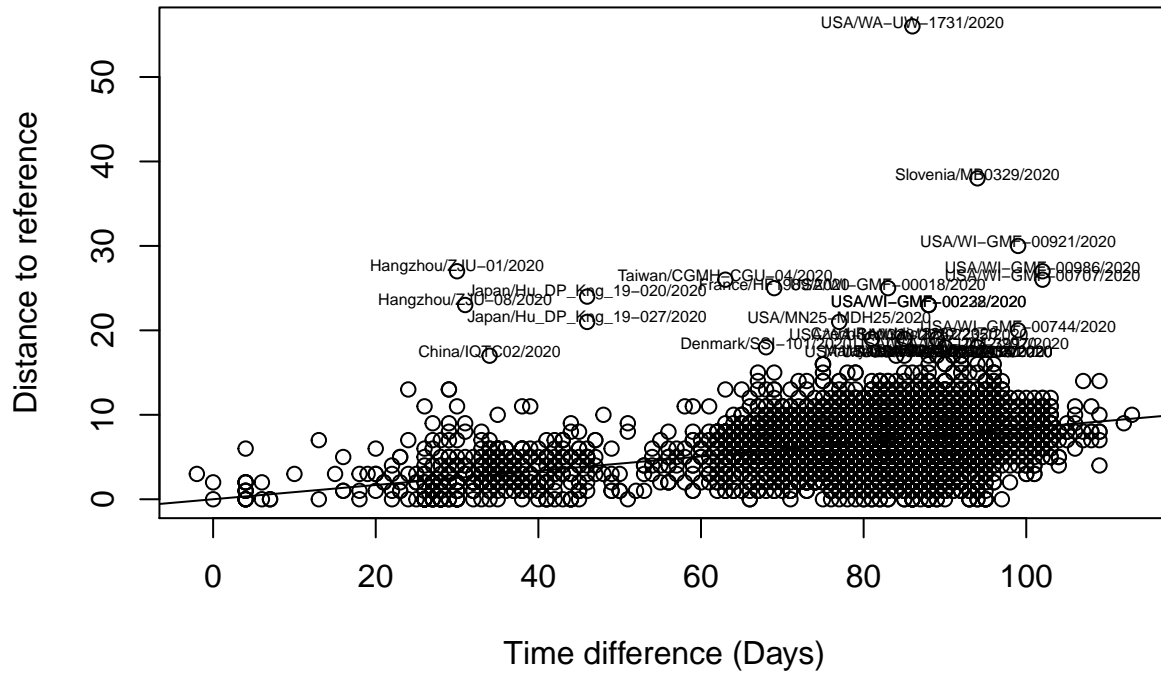
A rough estimate of the mutation rate can be obtained by regressing genetic divergence on time divergence. This estimate ignores correlation among samples, population structure and pairwise information. A more reliable estimate can be found in the mutation section.

```
##
## Call:
## lm(formula = hamming[ref, ] ~ -1 + x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.188 -1.415 -0.091  1.469 48.740
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## x 0.0844132   0.0003716   227.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.671 on 7663 degrees of freedom
## Multiple R-squared:  0.8707, Adjusted R-squared:  0.8707
## F-statistic: 5.161e+04 on 1 and 7663 DF, p-value: < 2.2e-16

##      China/IQTC02/2020|2020-01-29
##      1397
##      CzechRepublic/2282/2020|2020-03-20
##      1404
##      Denmark/SSI-101/2020|2020-03-03
##      1458
##      France/HF1989/2020|2020-03-04
##      2842
##      Hangzhou/ZJU-01/2020|2020-01-25
```

##		3116
##	Hangzhou/ZJU-08/2020 2020-01-26	
##		3125
##	Japan/Hu_DP_Kng_19-020/2020 2020-02-10	
##		3706
##	Japan/Hu_DP_Kng_19-027/2020 2020-02-10	
##		3707
##	Malaysia/189332/2020 2020-03-20	
##		3854
##	Slovenia/MB0329/2020 2020-03-29	
##		4969
##	Taiwan/CGMH-CGU-04/2020 2020-02-27	
##		5187
##	Taiwan/TSGH-10/2020 2020-03-24	
##		5232
##	USA/AZ-TG271435/2020 2020-03-27	
##		5305
##	USA/ID-UW-2255/2020 2020-03-23	
##		5447
##	USA/MN25-MDH25/2020 2020-03-12	
##		5540
##	USA/WA-UW-1588/2020 2020-03-19	
##		6482
##	USA/WA-UW-1731/2020 2020-03-21	
##		6530
##	USA/WA-UW-2142/2020 2020-03-25	
##		6637
##	USA/WA-UW-2235/2020 2020-03-24	
##		6660
##	USA/WA-UW-3841/2020 2020-03-26	
##		6685
##	USA/WA-UW-3935/2020 2020-03-27	
##		6707
##	USA/WA-UW-3954/2020 2020-03-27	
##		6714
##	USA/WA-UW-3997/2020 2020-03-29	
##		6727
##	USA/WA-UW330/2020 2020-03-16	
##		6969
##	USA/WI-GMF-00018/2020 2020-03-18	
##		7350
##	USA/WI-GMF-00228/2020 2020-03-23	
##		7351
##	USA/WI-GMF-00232/2020 2020-03-23	
##		7352
##	USA/WI-GMF-00707/2020 2020-04-06	
##		7359
##	USA/WI-GMF-00744/2020 2020-04-03	
##		7360
##	USA/WI-GMF-00921/2020 2020-04-03	
##		7361
##	USA/WI-GMF-00986/2020 2020-04-06	
##		7362

## A total of 2.36 of mutations per month

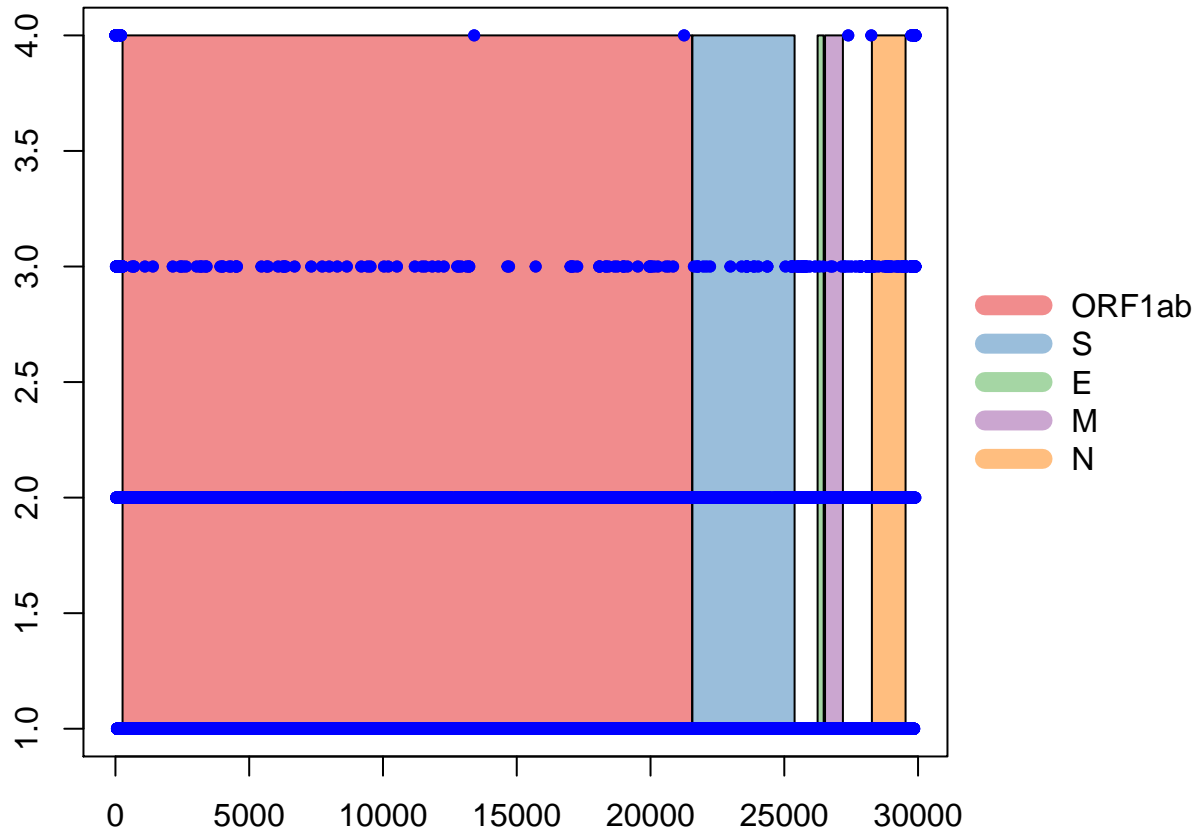


```
## [1] "China/IQTC02/2020"
## [1] "CzechRepublic/2282/2020"
## [1] "Denmark/SSI-101/2020"
## [1] "France/HF1989/2020"
## [1] "Hangzhou/ZJU-01/2020"
## [1] "Hangzhou/ZJU-08/2020"
## [1] "Japan/Hu_DP_Kng_19-020/2020"
## [1] "Japan/Hu_DP_Kng_19-027/2020"
## [1] "Malaysia/189332/2020"
## [1] "Slovenia/MB0329/2020"
## [1] "Taiwan/CGMH-CGU-04/2020"
## [1] "Taiwan/TSGH-10/2020"
## [1] "USA/AZ-TG271435/2020"
## [1] "USA/ID-UW-2255/2020"
## [1] "USA/MN25-MDH25/2020"
## [1] "USA/WA-UW-1588/2020"
## [1] "USA/WA-UW-1731/2020"
## [1] "USA/WA-UW-2142/2020"
## [1] "USA/WA-UW-2235/2020"
## [1] "USA/WA-UW-3841/2020"
## [1] "USA/WA-UW-3935/2020"
## [1] "USA/WA-UW-3954/2020"
## [1] "USA/WA-UW-3997/2020"
## [1] "USA/WA-UW330/2020"
## [1] "USA/WI-GMF-00018/2020"
## [1] "USA/WI-GMF-00228/2020"
## [1] "USA/WI-GMF-00232/2020"
## [1] "USA/WI-GMF-00707/2020"
## [1] "USA/WI-GMF-00744/2020"
```

```
## [1] "USA/WI-GMF-00921/2020"
## [1] "USA/WI-GMF-00986/2020"
```

To explore regions and number of polymorphic sites

```
## [1] 29903
```



Of the total 29903 sites in the reference sequence, 25691 are monomorphic sites, 3891 are binary, and only 321 have 3 or 4 types. Largest entropy is observed is in the first 700 bases and the last 26447 bases.

## Reference:

Rambaut, Andrew, Tommy T. Lam, Luiz Max Carvalho, and Oliver G. Pybus. "Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen)." *Virus evolution* 2, no. 1 (2016).