

Background

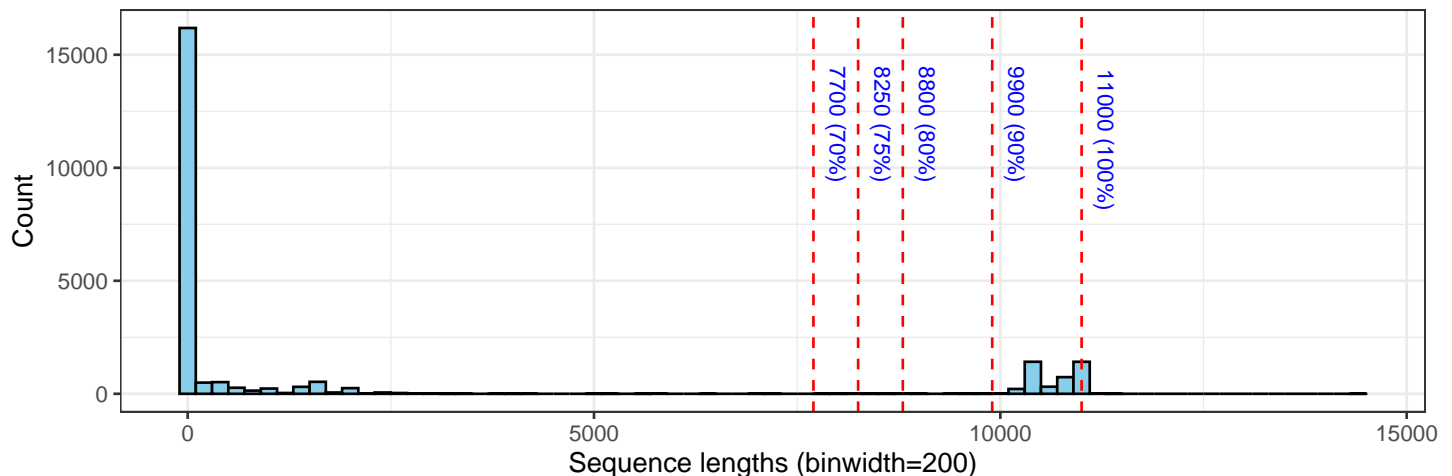
CDC website: <https://www.cdc.gov/west-nile-virus/index.html>

Exploratory Graphics

How long is the WNV genome?

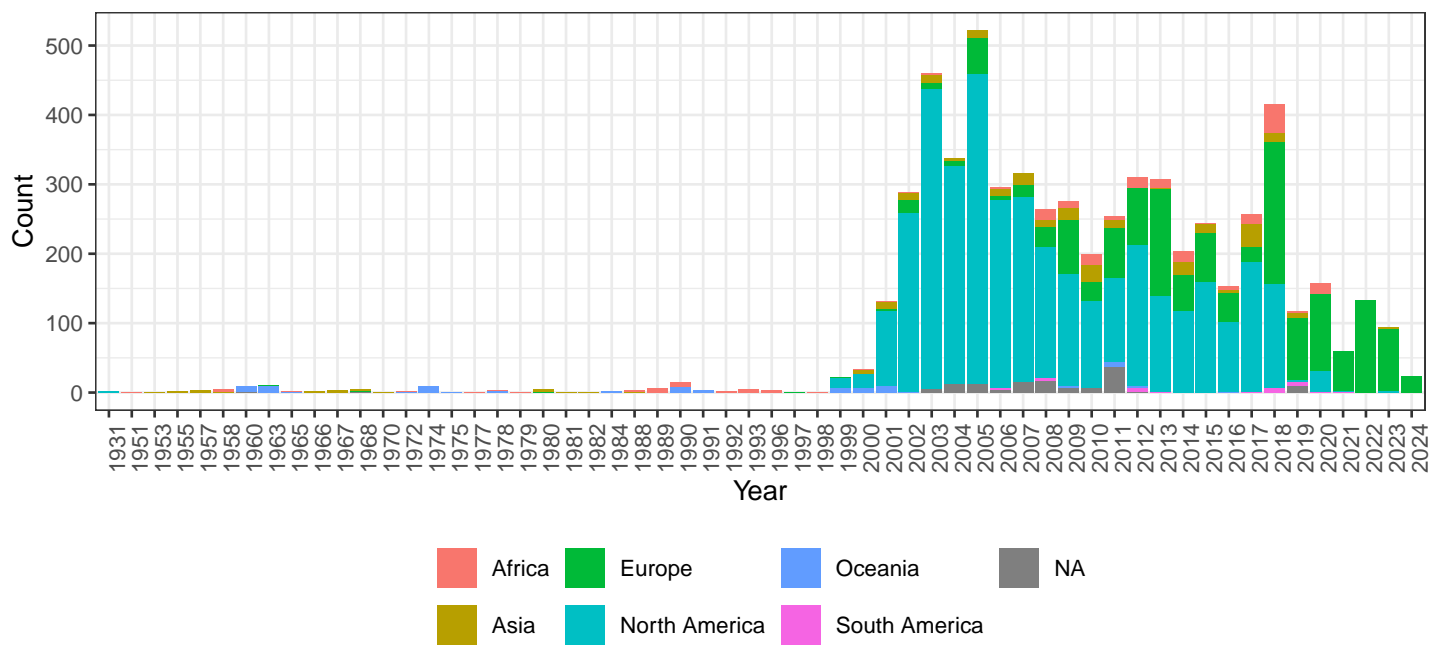
WNV is approximately 11k nt long. The samples pulled from NCBI GenBank have the following distribution of lengths. We can pick a length cut-off for the phylogenetic tree:

Sequence lengths (all data = 23336)



How well sampled is WNV across time and space?

WNV entries with collection date (n = 5992)



The earliest samples seem to be from 1931. However, these are EF631122 and EF631123 and are supposedly collected in Illinois, USA, much earlier than the NY99 event. We might need to dig into publications that contain these GenBanks to

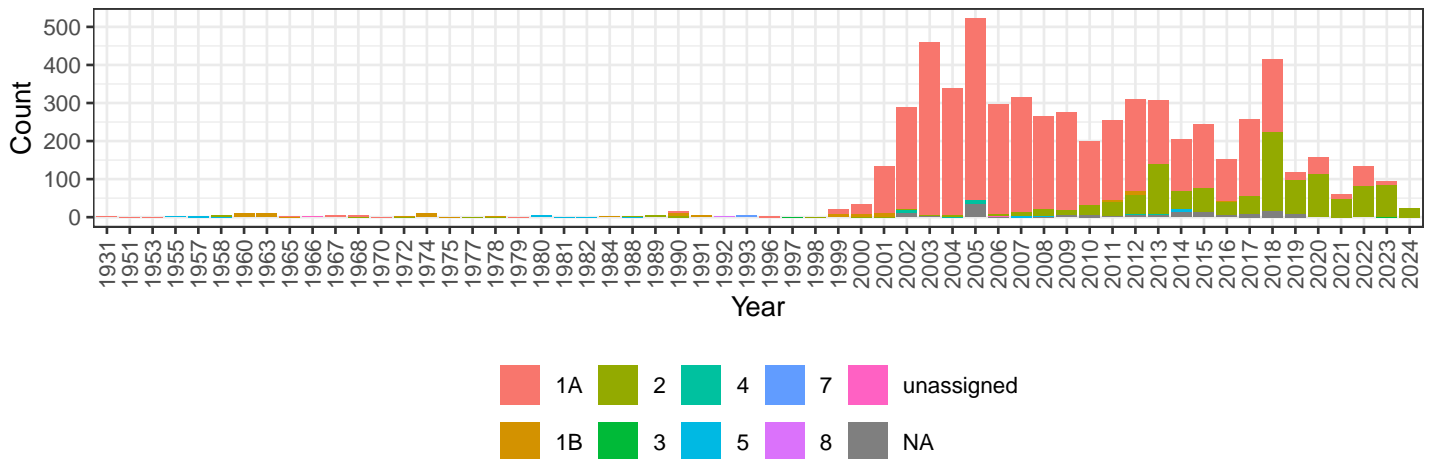
check if this collection date is correct. Such as publication:

- Amore, G., Bertolotti, L., Hamer, G.L., Kitron, U.D., Walker, E.D., Ruiz, M.O., Brawn, J.D. and Goldberg, T.L., 2010. Multi-year evolutionary dynamics of West Nile virus in suburban Chicago, USA, 2005–2007. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1548), pp.1871-1878.

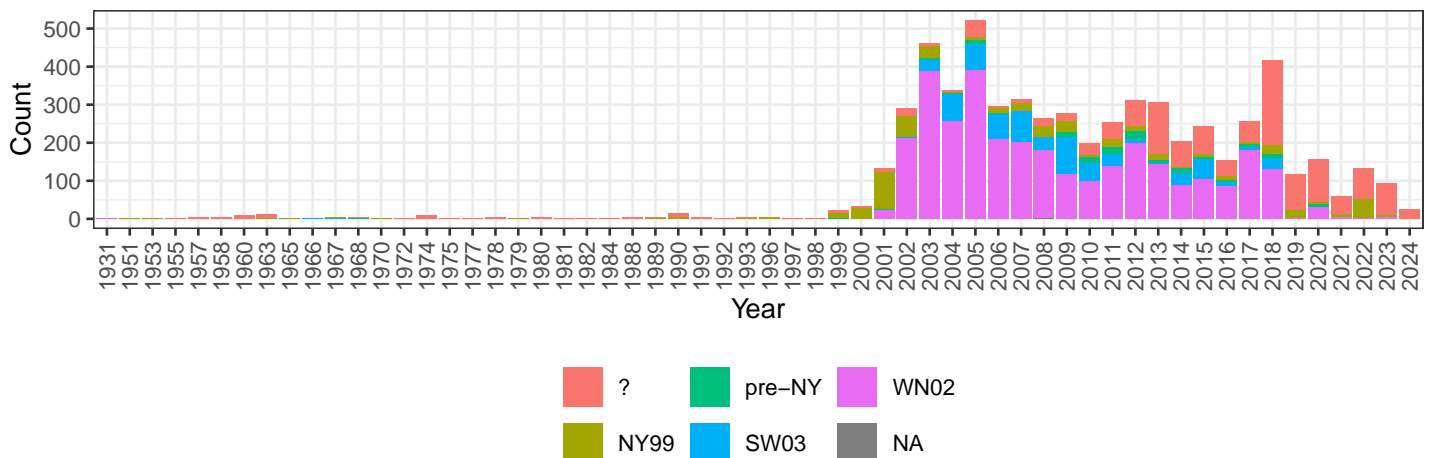
What lineages are available for WNV?

- **Global lineage system:** Koch, R. Tobias, Diana Erazo, Arran J. Folly, Nicholas Johnson, Simon Dellicour, Nathan D. Grubaugh, and Chantal BF Vogels. "Genomic epidemiology of West Nile virus in Europe." *One Health* (2023): 100664.
- **USA-based system:** Hadfield, J., Brito, A.F., Swetnam, D.M., Vogels, C.B., Tokarz, R.E., Andersen, K.G., Smith, R.C., Bedford, T. and Grubaugh, N.D., 2019. Twenty years of West Nile virus spread and evolution in the Americas visualized by Nextstrain. *PLoS pathogens*, 15(10), p.e1008042.

Count of global-based lineage samples with collection date (n = 5992)



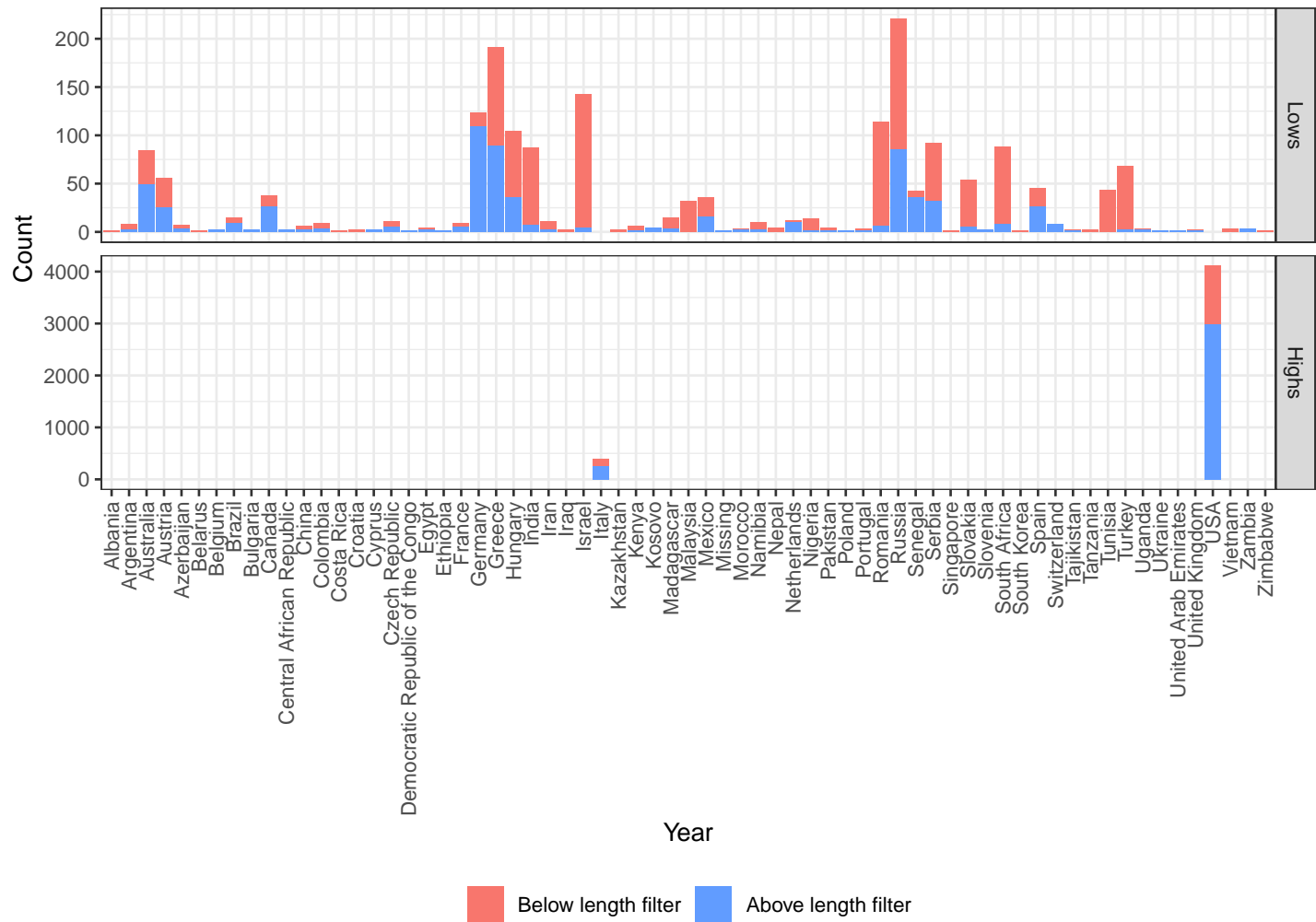
Count of USA-based lineage samples with collection date (n = 5992)



Global subsampling

If min length is set to 80 percent, or 8800nt:

WNV entries that were above or below filter, by country (n = 4147 where min-length=8800)



- Total WNV records = 23336. Note how this includes records without collection date
 - Total WNV records that passes the length filter = 4147
- If this number is lower than 4k, probably don't need to subsample further. Global build might need a different root... not the NY99