

# Homework Assignment

## Phylogenetic Trees

04-07-2023

### Task A

```
seqs <- read.fasta(file = "usflu.fasta")
```

### Task B

I just selected the first strain from each year.

```
annotations <- read.csv("usflu.annot.csv")
strains <- annotations %>%
  group_by(year) %>%
  slice_head(n = 1)
print(strains)
```

```
## # A tibble: 16 x 4
## # Groups:   year [16]
##       X accession year misc
##   <int> <chr>    <int> <chr>
## 1     1 CY013200  1993 (A/New York/783/1993(H3N2))
## 2     6 CY012272  1994 (A/New York/729/1994(H3N2))
## 3    11 CY012480  1995 (A/New York/666/1995(H3N2))
## 4    16 CY009476  1996 (A/New York/565/1996(H3N2))
## 5    21 CY006259  1997 (A/New York/511/1997(H3N2))
## 6    26 CY006787  1998 (A/New York/506/1998(H3N2))
## 7    31 CY001453  1999 (A/New York/184/1999(H3N2))
## 8    36 CY000737  2000 (A/New York/180/2000(H3N2))
## 9    41 CY002816  2001 (A/New York/301/2001(H3N2))
## 10   46 CY000297  2002 (A/New York/96/2002(H3N2))
## 11   51 CY000105  2003 (A/New York/60A/2003(H3N2))
## 12   56 CY019245  2004 (A/New York/908/2004(H3N2))
## 13   61 CY003640  2005 (A/New York/463/2005(H3N2))
## 14   66 EF554795  2006 (A/Ohio/2006(H3N2))
## 15   71 EU199369  2007 (A/Minnesota/08/2007(H3N2))
## 16   76 FJ549055  2008 (A/Illinois/14/2008(H3N2))
```

```
filtered_seqs <- seqs[strains$accession]
write.fasta(sequences = filtered_seqs, names = getName(filtered_seqs),
  file.out = "filtered_seqs.fasta")
```

## Task C

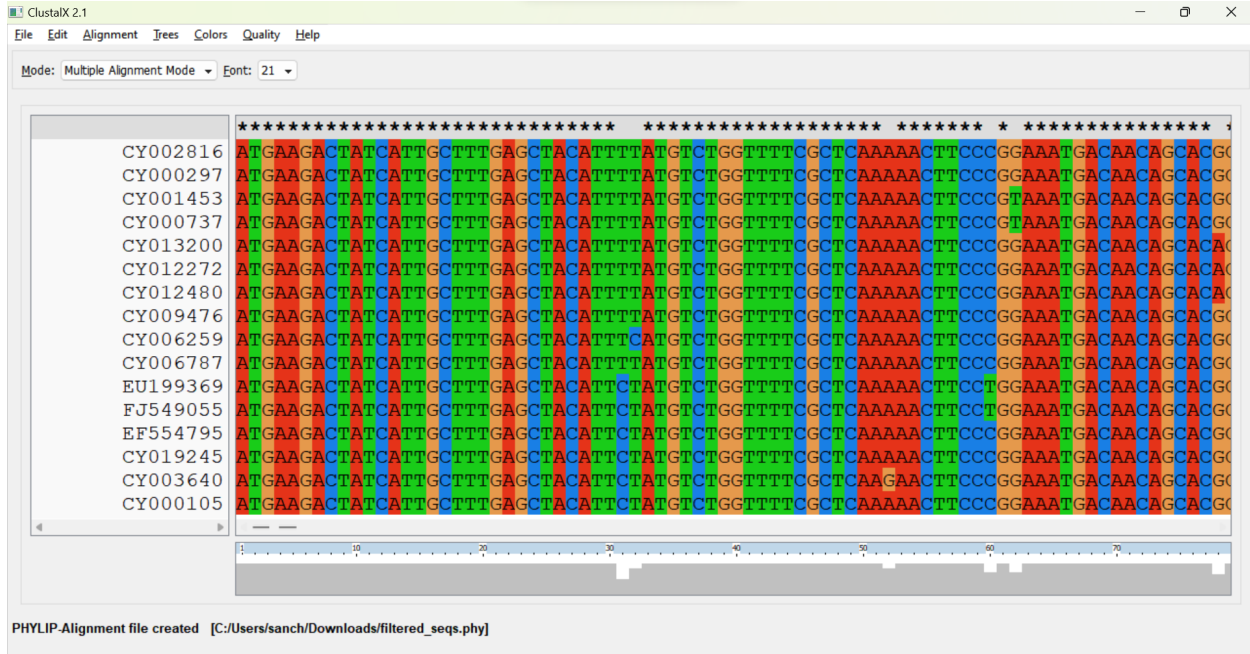


Figure 1: Alignment

## Task D