Exercise8_SequenceDifference

AUTHOR
Ruth Walters

PUBLISHED
February 17, 2026

```r
library(ape)
library(phangorn)
```

```r
ref <- c("U15717",
         "U15718",
         "U15719",
         "U15720",
         "U15721",
         "U15722",
         "U15723",
         "U15724")
```

```r
rampho_dna <- read.GenBank(ref)
rampho_dna
```

```
8 DNA sequences in binary format stored in a list.

All sequences of same length: 1045

Labels:
U15717
U15718
U15719
U15720
U15721
U15722
...

Base composition:
    a     c     g     t
0.267 0.351 0.134 0.247
(Total: 8.36 kb)
```

```r
raw <- dist.dna(rampho_dna,
                model = "raw",
                pairwise.deletion = TRUE)

raw
```

```
              U15717          U15718          U15719          U15720
U15721
U15718 0.0889952153
U15719 0.0497607656 0.0765550239
U15720 0.0153110048 0.0861244019 0.0468899522
U15721 0.0593301435 0.0746411483 0.0478468900 0.0526315789
U15722 0.0162679426 0.0851674641 0.0459330144 0.0009569378
0.0516746411
U15723 0.0660287081 0.0822966507 0.0612440191 0.0602870813
0.0382775120
U15724 0.0679425837 0.0784688995 0.0535885167 0.0583732057
0.0382775120
              U15722          U15723
U15718
U15719
U15720
U15721
U15722
U15723 0.0593301435
U15724 0.0574162679 0.0401913876
```

Question 1: The `dist.dna` function creates a matrix of pairwise distances between DNA sequences using a model of DNA evolution. So, the values represent the probability of one of the DNA sequences mutating into the other DNA sequence. The smaller the value, the higher the distance between sequences and the lower the probability that one becomes the other.

```
markov <- dist.dna(rampho_dna,
                   model = "JC69",
                   pairwise.deletion = TRUE)

plot(raw~markov,
     pch = 20,
     col = "red",
     xlim = c(0,.1),
     ylim = c(0, .1),
     xlab = "JC69 model distance",
     ylab = "Uncorrected distance",
     main = "Saturation plot")
abline(0, 1, lty = 2)
```
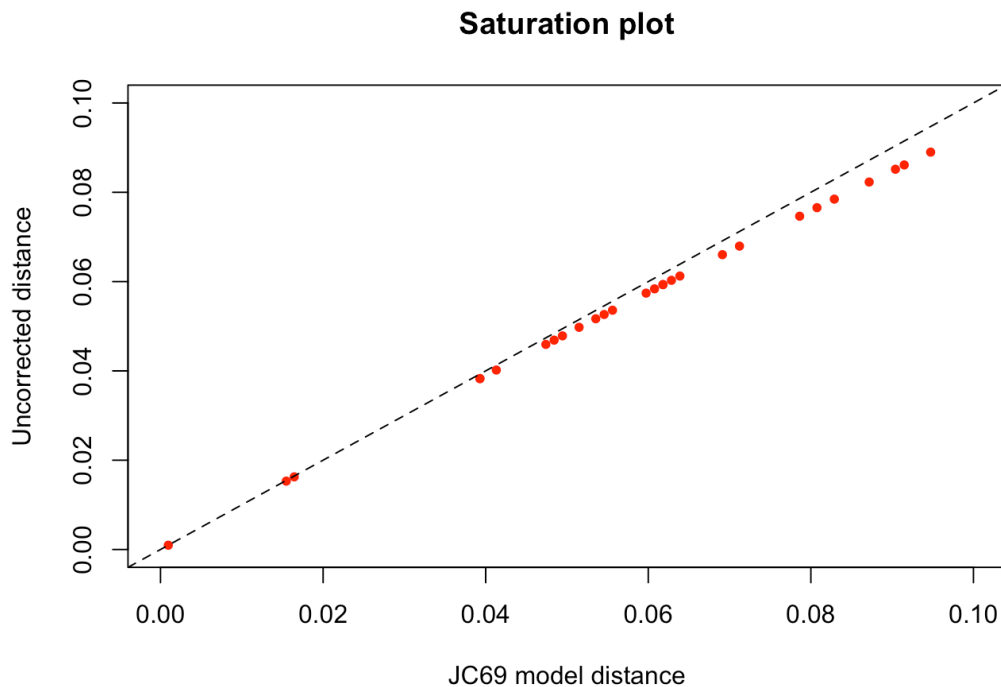
Saturation plot

Question 2: No; the data displayed does not significantly deviate from the line.

Question 3: Parsimony is prone to failure when the genes under study are approaching/have reached saturation. Under a parsimonious model, shared nucleotides are *always* reflective of inheritance. However, for more distantly related organisms, shared nucleotides might be the result of chance. This means that parsimony explains more closely related/nearer relationships in the tree better than more distant relations in the tree.

```
primates.dna <- read.dna("~/Documents/phylo/Exercise5/data/prima
                         format = "fasta")
primates.dna
```

```
7 DNA sequences in binary format stored in a matrix.

All sequences of same length: 1046

Labels:
ringtailed_lemur
howler_monkey
rhesus_monkey
gibbon
gorilla
humans
...
```
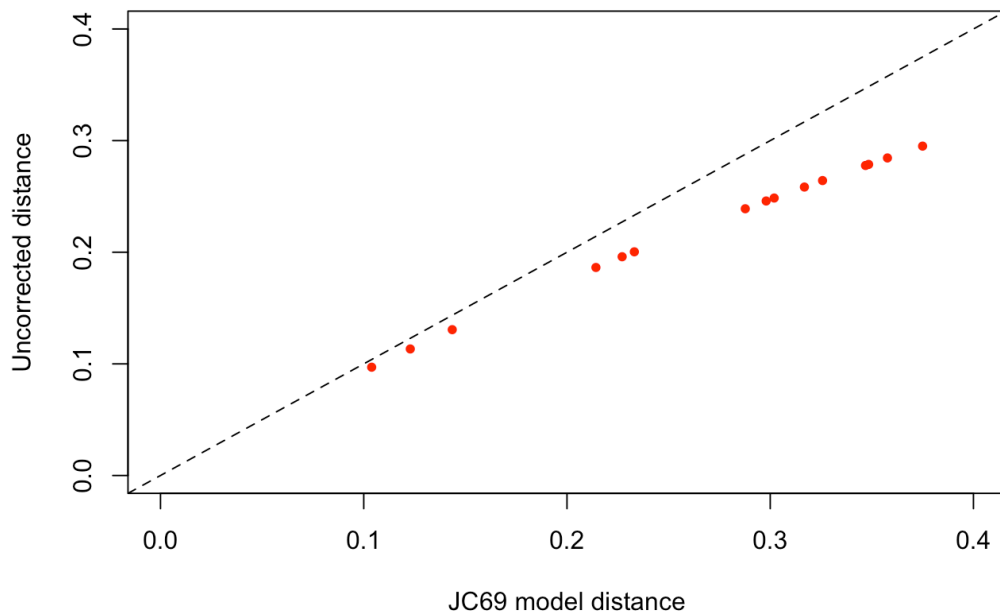
```
Base composition:
    a     c     g     t
0.329 0.321 0.088 0.261
(Total: 7.32 kb)
```

```r
raw <- dist.dna(primates.dna,
                model = "raw",
                pairwise.deletion = TRUE)

markov <- dist.dna(primates.dna,
                   model = "JC69",
                   pairwise.deletion = TRUE)

plot(raw~markov,
     pch = 20,
     col = "red",
     xlim = c(0,.4),
     ylim = c(0, .4),
     xlab = "JC69 model distance",
     ylab = "Uncorrected distance",
     main = "Saturation plot")
abline(0, 1, lty = 2)
```


Saturation plot

```r
primates.phyDat <- phyDat(primates.dna,
                          type = "DNA",
```
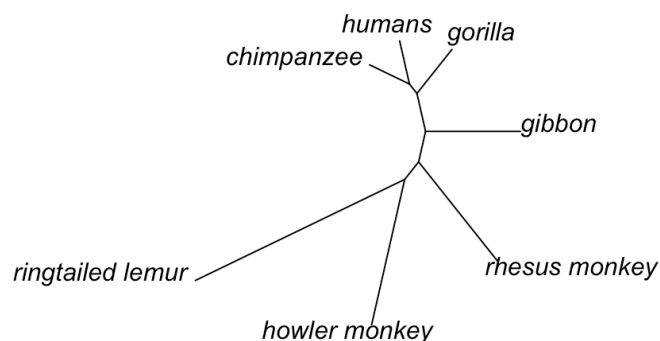
```
                    levels = NULL)

dm <- dist.dna(primates.dna,
               model = "JC69")

treeUPGMA <- upgma(dm)

plot(treeUPGMA,
     "unrooted",
     main = "JC69 corrected")
```

JC69 corrected

Question 5: The branches that separate ring-tailed lemurs and howler monkeys are long relative to some of the other branches in the tree.

Question 6: Given that the sequence for the primates DNA is saturated or approaching saturation, I suspect that a parsimonious approach would have the branches between these two organisms shorter. This reflects parsimony's tendency to view shared nucleotides as inherited nucleotides. I would not reccomend the use of parsimony in this instance.