

Computer Assignment 4

781552

10/15/2021

#Question 1 ## Subtask (a)

```
X <- c("C","A","G","C","C","C","T","A","G","T","T","G","C","C","C","C", "C","A","G","A","G","G","C","A")

states = c("GC", "AT")

T = cbind(c(0.80, 0.75), c(0.20, 0.25))
rownames(T) <- c("GC", "AT")
colnames(T) <- c("GC", "AT")

E = cbind(c(0.10, 0.40, 0.40, 0.10), c(0.40, 0.10, 0.10, 0.40))
rownames(E) <- c("A", "C", "G", "T")
colnames(E) <- c("GC", "AT")

hmm <- initHMM(States = states, Symbols = X, transProbs = T, emissionProbs = E)
bw = baumWelch(hmm, X)

bw

## $hmm
## $hmm$States
## [1] "GC" "AT"
##
## $hmm$Symbols
## [1] "C" "A" "G" "C" "C" "C" "T" "A" "G" "T" "T" "G" "C" "C" "C" "C" "C" "A" "G"
## [20] "A" "G" "G" "C" "A" "G" "G" "T" "A" "A" "A" "T" "A" "G" "C" "C" "A"
##
## $hmm$startProbs
##   GC  AT
## 0.5 0.5
##
## $hmm$transProbs
##      to
## from   GC      AT
##   GC 0.8260870 0.1739130
##   AT 0.4166667 0.5833333
##
## $hmm$emissionProbs
##      symbols
## states      C      A      G C C C      T A G T T G C C
##      GC 4.457141e-08 4.166666e-01 3.750000e-01 0 0 0 2.083333e-01 0 0 0 0 0 0
```

```
##      AT 1.000000e+00 6.086614e-10 1.857827e-14 0 0 0 7.745255e-36 0 0 0 0 0 0
##      symbols
## states C C C A G A G G C A G G T A A A T A G C C A
##      GC 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##      AT 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##
##
## $difference
## [1] 2.506104e+00 1.859027e-01 1.851036e-01 1.321911e-01 7.028590e-02
## [6] 3.574159e-02 2.644541e-02 2.198550e-02 1.776005e-02 1.421795e-02
## [11] 1.141354e-02 9.219188e-03 7.496765e-03 6.135352e-03 5.051088e-03
## [16] 4.181080e-03 3.477986e-03 2.905942e-03 2.437595e-03 2.051922e-03
## [21] 1.732648e-03 1.467074e-03 1.245212e-03 1.059150e-03 9.025752e-04
## [26] 7.704098e-04 6.585460e-04 5.636382e-04 4.829457e-04 4.142110e-04
## [31] 3.555657e-04 3.054558e-04 2.625842e-04 2.258638e-04 1.943805e-04
## [36] 1.673636e-04 1.441613e-04 1.242211e-04 1.070740e-04 9.232067e-05
## [41] 7.962077e-05 6.868382e-05 5.926146e-05 5.114120e-05 4.414095e-05
## [46] 3.810458e-05 3.289812e-05 2.840648e-05 2.453076e-05 2.118592e-05
## [51] 1.829878e-05 1.580636e-05 1.365441e-05 1.179622e-05 1.019151e-05
## [56] 8.805583e-06 7.608498e-06 6.574447e-06 5.681165e-06 4.909437e-06
## [61] 4.242685e-06 3.666599e-06 3.168826e-06 2.738702e-06 2.367017e-06
## [66] 2.045821e-06 1.768245e-06 1.528359e-06 1.321038e-06 1.141859e-06
## [71] 9.869961e-07 8.531477e-07 7.374597e-07 6.374662e-07 5.510367e-07
## [76] 4.763300e-07 4.117553e-07 3.559377e-07 3.076890e-07 2.659824e-07
## [81] 2.299305e-07 1.987663e-07 1.718270e-07 1.485396e-07 1.284089e-07
## [86] 1.110068e-07 9.596349e-08 8.295909e-08 7.171724e-08 6.199897e-08
## [91] 5.359774e-08 4.633508e-08 4.005662e-08 3.462899e-08 2.993685e-08
## [96] 2.588055e-08 2.237390e-08 1.934241e-08 1.672168e-08 1.445607e-08
```

```
bw$hmm$transProbs
```

```
##      to
## from      GC      AT
## GC 0.8260870 0.1739130
## AT 0.4166667 0.5833333
```

```
bw$hmm$emissionProbs
```

```
##      symbols
## states      C      A      G C C C      T A G T T G C C
## GC 4.457141e-08 4.166666e-01 3.750000e-01 0 0 0 2.083333e-01 0 0 0 0 0 0
## AT 1.000000e+00 6.086614e-10 1.857827e-14 0 0 0 7.745255e-36 0 0 0 0 0 0
##      symbols
## states C C C A G A G G C A G G T A A A T A G C C A
## GC 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## AT 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

```
##Subtask (b)
```

```
newhmm <- initHMM(States = states, Symbols = X, transProbs = bw$hmm$transProbs , emissionProbs = bw$hmm$emissionProbs)
viterbi(newhmm, X)
```

```
## [1] "AT" "GC" "GC" "AT" "AT" "AT" "GC" "GC" "GC" "GC" "GC" "GC" "AT" "AT" "AT"
## [16] "AT" "AT" "GC" "GC" "GC" "GC" "GC" "AT" "GC" "GC" "GC" "GC" "GC" "GC" "GC"
## [31] "GC" "GC" "GC" "AT" "AT" "GC"
```

##Subtask (c)

```
fw = forward(newhmm, X)
bw = backward(newhmm, X)

fw
```

```
##      index
## states      1      2      3      4      5      6
##      GC -17.6193204 -2.444085 -3.615969 -20.733198 -23.166811 -23.705807
##      AT -0.6931472 -22.451903 -35.810068 -5.365169 -5.904166 -6.443162
##      index
## states      7      8      9     10     11     12     13
##      GC -8.887247 -9.953771 -11.12566 -12.88533 -14.64500 -15.81688 -32.93411
##      AT -87.828141 -31.856206 -43.31975 -93.72084 -95.48051 -48.01098 -17.56608
##      index
## states     14     15     16     17     18     19     20
##      GC -35.36772 -35.90672 -36.44572 -36.98471 -21.47301 -22.64489 -23.71141
##      AT -18.10508 -18.64407 -19.18307 -19.72207 -41.48082 -54.83899 -45.61385
##      index
## states     21     22     23     24     25     26     27
##      GC -24.8833 -26.05518 -43.17241 -29.55532 -30.7272 -31.89909 -33.65876
##      AT -57.0774 -58.24928 -27.80438 -49.56314 -62.9213 -64.09319 -114.49427
##      index
## states     28     29     30     31     32     33     34
##      GC -34.72528 -35.79181 -36.85833 -38.6180 -39.68453 -40.85641 -57.97364
##      AT -56.62772 -57.69424 -58.76077 -119.4535 -61.58696 -73.05051 -42.60561
##      index
## states     35     36
##      GC -60.40725 -44.89555
##      AT -43.14461 -64.90336
```

```
bw
```

```
##      index
## states      1      2      3      4      5      6      7
##      GC -43.51799 -42.45146 -41.27958 -40.74058 -40.20158 -37.76797 -36.00830
##      AT -44.20240 -43.13587 -40.06937 -39.53038 -38.99138 -38.45238 -36.69271
##      index
## states      8      9     10     11     12     13     14
##      GC -34.94177 -33.76989 -32.01022 -30.25055 -29.07866 -28.53967 -28.00067
##      AT -35.62619 -34.45430 -32.69463 -30.93496 -27.86846 -27.32946 -26.79047
##      index
## states     15     16     17     18     19     20     21
##      GC -27.46167 -26.92268 -24.48906 -23.42254 -22.25066 -21.18413 -20.01225
##      AT -26.25147 -25.71247 -25.17348 -24.10695 -22.93507 -21.86855 -20.69666
##      index
## states     22     23     24     25     26     27     28
##      GC -18.84036 -16.40675 -15.34023 -14.16834 -12.99646 -11.23679 -10.17026
```

```
##      AT -17.63016 -17.09116 -16.02464 -14.85275 -13.68087 -11.92120 -10.85468
##      index
## states      29      30      31      32      33      34      35 36
##      GC -9.103738 -8.037214 -6.277542 -5.211018 -4.039134 -3.500137 -1.066524 0
##      AT -9.788151 -8.721627 -6.961956 -5.895432 -2.828930 -2.289934 -1.750938 0
```

##Subtask d

```
probSeq = fw[1, 36] + fw[2, 36]
probSeq
```

```
## [1] -109.7989
```

##Subtask e

```
posterior(newhmm, X)
```

```
##      index
## states      1      2      3      4      5
##      GC 8.836766e-08 1.000000e+00 1.000000e+00 6.311976e-08 9.491698e-09
##      AT 9.999999e-01 1.031521e-09 3.498365e-14 9.999999e-01 1.000000e+00
##      index
## states      6      7      8      9     10
##      GC 6.311976e-08 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00
##      AT 9.999999e-01 2.625234e-35 1.551159e-10 5.260698e-15 3.947719e-36
##      index
## states     11     12     13     14     15
##      GC 1.000000e+00 1.000000e+00 6.311976e-08 9.491695e-09 9.491693e-09
##      AT 3.947719e-36 3.498365e-14 9.999999e-01 1.000000e+00 1.000000e+00
##      index
## states     16     17     18     19     20
##      GC 9.491695e-09 6.311976e-08 1.000000e+00 1.000000e+00 1.000000e+00
##      AT 1.000000e+00 9.999999e-01 1.031521e-09 5.260698e-15 1.551159e-10
##      index
## states     21     22     23     24     25
##      GC 1.000000e+00 1.000000e+00 4.197463e-07 1.000000e+00 1.000000e+00
##      AT 5.260698e-15 3.498364e-14 9.999996e-01 1.031521e-09 5.260698e-15
##      index
## states     26     27     28     29     30
##      GC 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00
##      AT 5.260698e-15 3.947719e-36 1.551159e-10 1.551159e-10 1.551159e-10
##      index
## states     31     32     33     34     35
##      GC 1.000000e+00 1.000000e+00 1.000000e+00 6.311978e-08 6.311978e-08
##      AT 3.947719e-36 1.551159e-10 3.498365e-14 9.999999e-01 9.999999e-01
##      index
## states     36
##      GC 1.000000e+00
##      AT 2.045102e-09
```

##Subtask f

The results from the posterior allow us to see the path the sequence will take

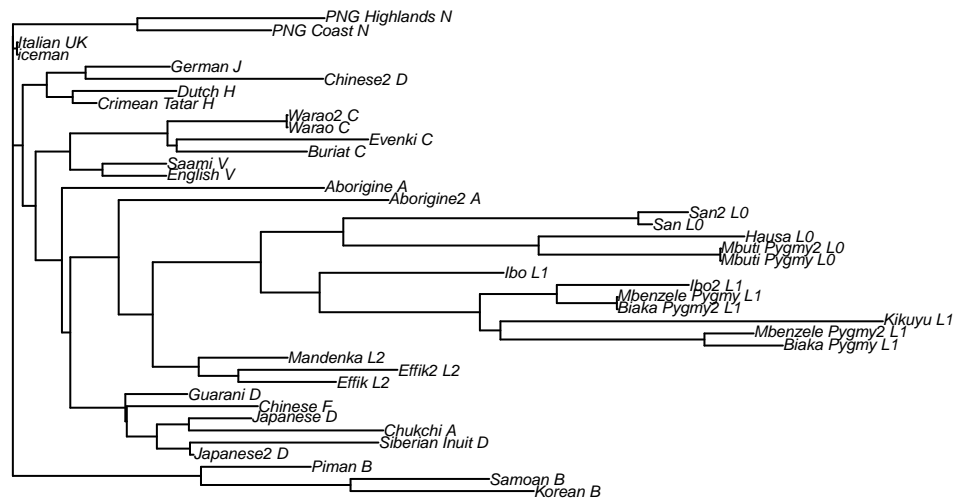
#Question 2 ##Subtask a

```
load("D:/Downloads/iceMan_Phylo.RData")
x <- as.DNABin(iceMan_phyDat)

dist_a <- dist.dna(x)

tree_a <- NJ(dist_a)

plot(tree_a, cex = 0.5)
```

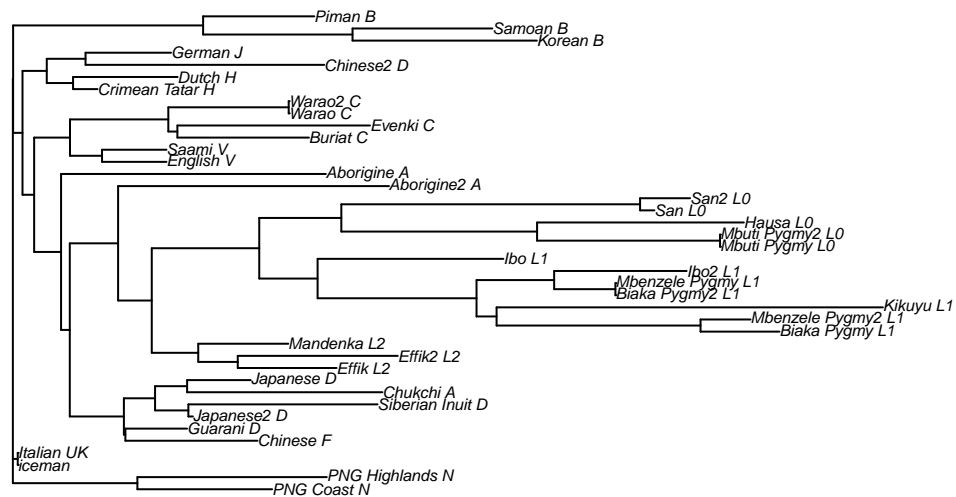


##Subtask b

```
dist_b <- dist.dna(x, model = "JC69")

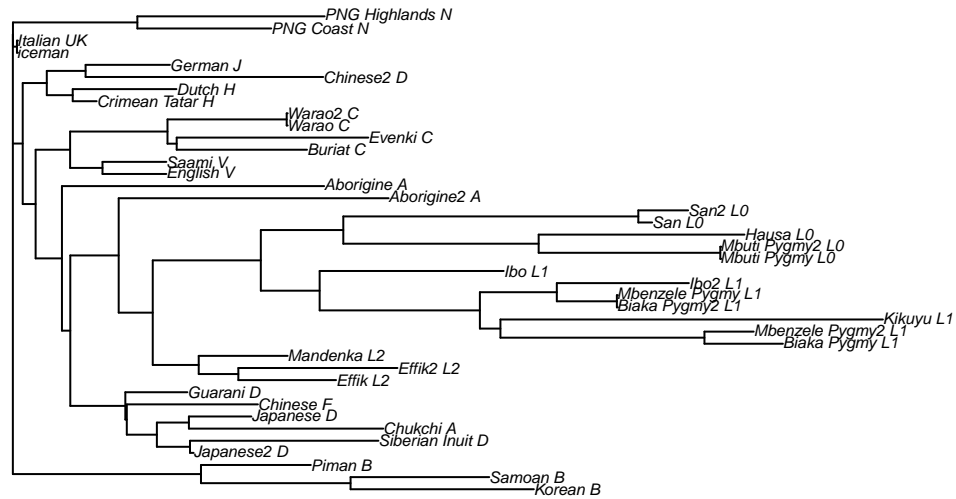
tree_b <- NJ(dist_b)

plot(tree_b, cex = 0.5)
```



##Subtask c

```
dist_c <- dist.dna(x, model = "K80")
tree_c <- NJ(dist_c)
plot(tree_c, cex = 0.5)
```



The most related modern human population to the Iceman is the Italian UK

The trees all differ because of the different algorithms used to calculate them.

3.

a.

Homo sapiens BRCA1 DNA repair associated (BRCA1), transcript variant 6, n

Sequence ID: [NR_027676.2](#) Length: 7152 Number of Matches: 1

Range 1: 87 to 106 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

| Score | Expect | Identities | Gaps | Strand |
|---------------|--------|----------------------|----------|------------|
| 37.4 bits(40) | 1.0 | 20/20(100%) | 0/20(0%) | Plus/Minus |
| Query | 57 | GGACTCTACTACCTTTACCC | 76 | |
| | | | | |
| Sbjct | 106 | GGACTCTACTACCTTTACCC | 87 | |

Homologous sequence :

```
Query  57      GGACTCTACTACCTTTACCC  76
          |||||
Sbjct  106      GGACTCTACTACCTTTACCC  87
```

Gene name:

NR_027676.2

Location: 87-106

Significance rate: 100%

b.

BRAF protein, partial [Homo sapiens]

Sequence ID: [ADQ00186.1](#) Length: 39 Number of Matches: 1

[See 4 more title\(s\)](#) ▼ [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 39 [GenPept](#) [Graphics](#)

▼ [Next Match](#) ▲ [Previous Match](#)

| Score | Expect | Method | Identities | Positives | Gaps |
|----------------|---|------------------------------|-------------|-------------|----------|
| 80.9 bits(198) | 5e-22 | Compositional matrix adjust. | 39/39(100%) | 39/39(100%) | 0/39(0%) |
| Query 1 | IFLHEDLTVKIGDFGLATVKSRWSGSHQFEQLSGSILWM | | 39 | | |
| | IFLHEDLTVKIGDFGLATVKSRWSGSHQFEQLSGSILWM | | | | |
| Sbjct 1 | IFLHEDLTVKIGDFGLATVKSRWSGSHQFEQLSGSILWM | | 39 | | |

Function: The B-Raf protein is involved in sending signals inside cells which are involved in directing cell growth

Accession number: ADQ00186.1

E-value: $5e^{-22}$