Pset4

1a.

i)

Log likelihood for topology 1 is -2746.80

Log likelihood for topology 2 is -2923.72

Log likelihood for topology 3 is -2923.72

# Topology 1 is the maximum likelihood tree

ii)

Amplify branch length for branch 1-4 by 1000*
thus, branch length will be:

```
branch_len = matrix(c(0.07517*1000, 0.03059*1000, 0.03161*1000, 0.11761*1000, 0.14289,
                      0.20843*1000, 0.03397*1000, 0.03497*1000, 0.24952*1000, 0.00000,
                      0.20843*1000, 0.03397*1000, 0.03497*1000, 0.24952*1000, 0.00000),ncol = 5, nrow = 3,byrow = TRUE)
```

Outcomes of the modified branch length:

Log likelihood for topology 1 is -5140.38

Log likelihood for topology 2 is -5140.38

Log likelihood for topology 3 is -5140.38

The log likelihood of the three trees are exactly the same, making it less likely to distinguish the best tree.
As the branch 5 getting shorter, the difference between the trees are smaller and it is harder to find the maximum likelihood tree.

iii)

Or we can test in some other genes/longer coding region except for APOE. The optimal tree topology should be consistent with the outcomes of APOE.
For the fixed species, we can also test with Sankoff's algrothm by minimizing the base transitions in each node and get the best tree topology.

1b. output:

root:

ATGAAGGTTCTGTGGGCTGCGCTGCTGGTCACATTCCTGGCAGGATGCCAGGCCGAGGT

GGAGCTGGAGA

human:

ATGAAGGTTCTGTGGGCTGCGTTGCTGGTCACATTCCTGGCAGGATGCCAGGCCAAGGTG
GAGCTGGAGA

mouse:

ATGAAGGCTCTGTGGGCCGTGCTGTTGGTCACATTGCTGACAGGATGCCTAGCCGAGGGA
GAGCAGGTGA

rat:

ATGAAGGCTCTGTGGGCCCTGCTGTTGGTCCCATTGCTGACAGGATGCCTGGCCGAGGGA
GAGCAGGTGA

dog:

ATGAAGGTTCTGTGGGCTGCGCTGGTGGTCACGCTCCTGGCAGGATGCTGGGCCGATGT
GCAGCCGGAGC

root:

CAGAGCAGACCGAGTGGCAGAGCGGCCAGCCCTGGGAGCTGGCACTGGGCCGCTTCTG
GGATTACCTGCG

human:

CAGAGCAGACCGAGTGGCAGAGCGGCCAGCGCTGGGAACTGGCACTGGGTCGCTTTTGG
GATTACCTGCG

mouse:

CAGATCAGCTCGAGTGGCAAAGCAACCAACCCTGGGAGCAGGCCCTGAACCGCTTCTGG
GATTACCTGCG

rat:

CAGATCAGCTCCCAGGGCAAAGCGACCAACCCTGGGAGCAGGCCCTGAACCGCTTCTGG
GATTACCTGCG

dog:

CGGAGCAGGCCGGGTGGCAGACTGGCCAGCCCTGGGAGGCGGCGCTGGCCCGCTTCTG
GGATTACCTGCG

root:

CTGGGTGCAGACGCTGTCTGACCAGGTGCAGGAGGAGCTGCTCAGCTCCCAGGTCACCC
AGGAACTGACG

human:

CTGGGTGCAGACACTGTCTGAGCAGGTGCAGGAGGAGCTGCTCAGCTCCCAGGTCACCC
AGGAACTGAGG

mouse:

CTGGGTGCAGACGCTGTCTGACCAGGTCCAGGAAGAGCTGCAGAGCTCCCAAGTCACACA
AGAACTGACG

rat:

CTGGGTGCAGACGCTTTCTGACCAGGTCCAGGAAGAGCTGCAGAGCTCCCAAGTCACACA
GGAACTGACG

dog:

CTGGGTGCAGACGCTGTCTGACCAGGTGCAAGAGGGCGTGCTCAACACCCAGGTCACCC
AGGAACTGACG

root:

GCGCTGATGGACGAGACCATGAAGGAGGTGAAGGCCTACAAATCGGAGCTGGAGGAACA
GCTGGGCCCGG

human:

GCGCTGATGGACGAGACCATGAAGGAGTTGAAGGCCTACAAATCGGAACTGGAGGAACAA
CTGACCCCGG

mouse:

GCACTGATGGAGGACACTATGACGGAAGTAAAGGCTTACAAAAAGGAGCTGGAGGAACAG
CTGGGTCCAG

rat:

GTACTGATGGAGGACACTATGACGGAAGTAAAGGCATACAAAAAGGAGCTGGAGGAACAG
CTGGGCCCAG

dog:

GCGCTGATGGATGAGACCATGAAGGAGGTGAAGGCCTACAAGGCGGAGCTGGACGAGCA
GCTGGGCCCCA

root:

TGGCGGAGGAGACGCGGGCCCGGCTGGCCAAGGAGCTGCAGGCGGCGCAGGCCCGGC
TGGGCGCGGACAT

human:

TGGCGGAGGAGACGCGGGCACGGCTGTCCAAGGAGCTGCAGGCGGCGCAGGCCCGGCT
GGGCGCGGACAT

mouse:

TGGCGGAGGAGACACGGGCCAGGCTGGGCAAAGAGGTGCAGGCGGCACAGGCCCGACT
CGGAGCCGACAT

rat:

TGGCGGAGGAGACACGGGCCAGGCTGGCTAAAGAGGTGCAGGCGGCACAGGCCCGTCT
GGGAGCTGACAT

dog:

TGACCTCGGAGACGCAGGCCCGCGTGGCCAAGGAGCTGCAGGCGGCGCAGGCCCGGCT
GCGCTCGGACAT

root:

GGAGGACGTGCGCAACCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTGGGC
CAGAGCACCGAG

human:

GGAGGACGTGTGCGGCCGCCTGGTGCAGTACCGCGGCGAGGTGCAGGCCATGCTCGGC
CAGAGCACCGAG

mouse:

GGAGGATCTACGCAACCGACTCGGGCAGTACCGCAACGAGGTGCACACCATGCTGGGCC
AGAGCACAGAG

rat:

GGAGGATCTACGCAACCGACTCGGGCAGTACCGCAACGAGGTAAACACCATGCTGGGCC
AGAGCACAGAG

dog:

GGAGGACGTGCGCAACCGCCTGACGCAGTACCGCGGCGAGCTGCAGGCCATGCTGGGC
CAGAGCAGCGAG

root:

GAGCTGCGGGCGCGCCTCGCCTCCCACCTGCGCAAGCTGCGCAAGCGGCTGCTGCGGG
ATGCCGATGACC

human:

GAGCTGCGGGTGCGCCTCGCCTCCCACCTGCGCAAGCTGCGTAAGCGGCTCCTCCGCGA
TGCCGATGACC

mouse:

GAGATACGGGCGCGGCTCTCCACACACCTGCGCAAGATGCGCAAGCGCTTGATGCGGGA
TGCCGAGGATC

rat:

GAGCTGCGGTCGCGCCTCTCCACACACCTGCGCAAGATGCGCAAGCGCCTGATGCGGGA
TGCGGATGATC

dog:

GAGCTGCGGGCGCGCTTCGCCTCCCACATGCGCAAGCTGCGCAAGCGGGTGCTGCGGG
ACGCCGAGGACC

root:

TGCAGAAGCGCCTGGCAGTGTACAAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCGT
GAGCGCCATCCG

human:

TGCAGAAGCGCCTGGCAGTGTACCAGGCCGGGGCCCGCGAGGGCGCCGAGCGCGGCCT
CAGCGCCATCCG

mouse:

TGCAGAAGCGCCTAGCTGTGTACAAGGCAGGGGCACGCGAGGGCGCCGAGCGCGGTGT
GAGTGCCATCCG

rat:

TGCAGAAGCGCCTGGCGGTGTACAAGGCCGGGGCACAGGAGGGCGCCGAGCGCGGTGT
GAGTGCTATCCG

dog:

TGCAGAGGCGCCTGGCCGTCTACAAGGCCGGCGTGCGCGAGGGTGCCGAGCGCAGCGT
GAGCAGCATCCG

root:

CGAGCGCCTGGGGCCCCTGGTGGAGCAGGGCCGCGAGCGCACCGCCAATGTGGGCGCC
CTGGCCGCCCAG

human:

CGAGCGCCTGGGGCCCCTGGTGGAACAGGGCCGCGTGCGGGCCGCCACTGTGGGCTCC
CTGGCCGGCCAG

mouse:

TGAGCGCCTGGGGCCTCTGGTGGAGCAAGGTCGCCAGCGCACTGCCAACCTAGGCGCTG
GGGCCGCCCAG

rat:

TGAGCGCCTGGGGCCACTGGTGGAGCAGGGTCGTCAGCGCACAGCCAACCTAGGCGCTG
GCGCCGCCCAG

dog:

CGAGCGCCTCTGGCCGCTGCTGGAGCAGGCCCGCGAGCGCAACGCCAAGGTGGGCGCC
CTGGCCACGCAG

root:

CCGCTGCACGAGCGGGCCCAGGCCTTGGGCGAGCGGCTGCGCGGGCGGCTGGAGGAGA
TGGGCAGCCGGG

human:

CCGCTACAGGAGCGGGCCCAGGCCTGGGGCGAGCGGCTGCGCGCGCGGATGGAGGAGA
TGGGCAGCCGGA

mouse:

CCTCTGCGCGATCGCGCCCAGGCTTTTGGTGACCGCATCCGAGGGCGGCTGGAGGAAGT
GGGCAACCAGG

rat:

CCCCTGCGCGATCGCGCCCAGGCTTTGAGTGACCGCATCCGAGGGCGGCTGGAGGAAGT
GGGCAACCAGG

dog:

CCGCTGCTCGAGCGGGCCGACGCCCTGGGCCAGCAGCTGCGCGGGCAGCTGGAGGAGA
TGAGCAGCCGGG

root:

CCCGCGACCGCCTGGAGGAGGTGCGGGAGCAGATGGAGGAGGTGCGCGCCAAGATGGA
GGAGCAGGCCCA

human:

CCCGCGACCGCCTGGACGAGGTGAAGGAGCAGGTGGCGGAGGTGCGCGCCAAGCTGGA
GGAGCAGGCCCA

mouse:

CCCGTGACCGCCTAGAGGAGGTGCGTGAGCACATGGAGGAGGTGCGCTCCAAGATGGAG
GAACAGACCCA

rat:

CCCGAGACCGCCTAGAGGAGGTGCGTGAGCAGATGGAGGAGGTGCGCTCCAAGATGGAG
GAGCAGACCCA

dog:

CCCGCGGCCACCTGGAGGAGATGCGCGAGCAGATACAGGAGGTGCGGGTGAAGATGGA
GGAGCAGGCCGA

root:

GCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGAGCTGGTTCGAGCCCC
TGGTGGAAGAC

human:

GCAGATACGCCTGCAGGCCGAGGCCTTCCAGGCCCGCCTCAAGAGCTGGTTCGAGCCCC
TGGTGGAAGAC

mouse:

GCAAATACGCCTGCAGGCGGAGATCTTCCAGGCCCGCCTCAAGGGCTGGTTCGAGCCAAT
AGTGGAAGAC

rat:

GCAGATACGCCTGCAGGCCGAGATCTTCCAGGCCCGCATCAAGGGCTGGTTCGAGCCGC
TAGTGGAAGAC

dog:

CCAGATACGCCAAAAGGCCGAGGCCTTCCAGGCGCGCCTCAAGAGCTGGTTCGAGCCCC
TGCTGGAAGAC

root:

ATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGCCACCAGCC
CCACCCCTGTGC

human:

ATGCAGCGCCAGTGGGCCGGGCTGGTGGAGAAGGTGCAGGCTGCCGTGGGCACCAGCG
CCGCCCCTGTGC

mouse:

ATGCATCGCCAGTGGGCAAACCTGATGGAGAAGATACAGGCCTCTGTGGCTACCAACCCC
ATCCCAGTGG

rat:

ATGCAGCGCCAGTGGGCAAACCTAATGGAGAAGATACAGGCCTCTGTGGCTACCAACTCC
ATCACAGTGC

dog:

ATGCAGCGCCAGTGGGACGGGCTGGTGGAGAAGGTGCAGGCGGCCGTGGCCACCATCC
CCACCTCTAAGC

root: CCATGGAGAATCAATGA

human: CCAGCGACAATCACTGA

mouse: CCCAGGAGAATCAATGA

rat: CCCTGGAGAATCAATGA

dog: CTGTGGAGGAACCATGA

1c.

Note: I highlighted the most of the difference by hand…

root:

MKVLWAALLVTFLAGCQAEVELETEQTEWQSGQPWELALGRFWDYLRWVQTLSDQVQEELL
SSQVTQELT

human:

MKVLWAALLVTFLAGCQAKVELETEQTEWQSGQRWELALGRFWDYLRWVQTLSEQVQEELL
SSQVTQELR

root:

ALMDETMKEVKAYKSELEEQLGPVAEETRARLAKELQAAQARLGADMEDVRNRLVQYRGEVQ
AMLGQSTE

human:

ALMDETMKELKAYKSELEEQLTPVAEETRARLSKELQAAQARLGADMEDVCGRLVQYRGEVQ
AMLGQSTE

root:

ELRARLASHLRKLRKRLLRDADDLQKRLAVYKAGAREGAERGVSAIRERLGPLVEQGRERTAN
VGALAAQ

human:

ELRVRLASHLRKLRKRLLRDADDLQKRLAVYQAGAREGAERGLSAIRERLGPLVEQGRVRAAT
VGSLAGQ

root:

PLHERAQALGERLRGRLEEMGSRARDRLEEVREQMEEVRAKMEEQAQQIRLQAEAFQARLKS
WFEPLVED

human:

PLQERAQAWGERLRARMEEMGSRTRDRLDEVKEQVAEVRAKLEEQAQQIRLQAEAFQARLK
SWFEPLVED

root:  MQRQWAGLVEKVQAAVATSPTPVPMENQ$

human: MQRQWAGLVEKVQAAVGTSAAPVPSDNH$

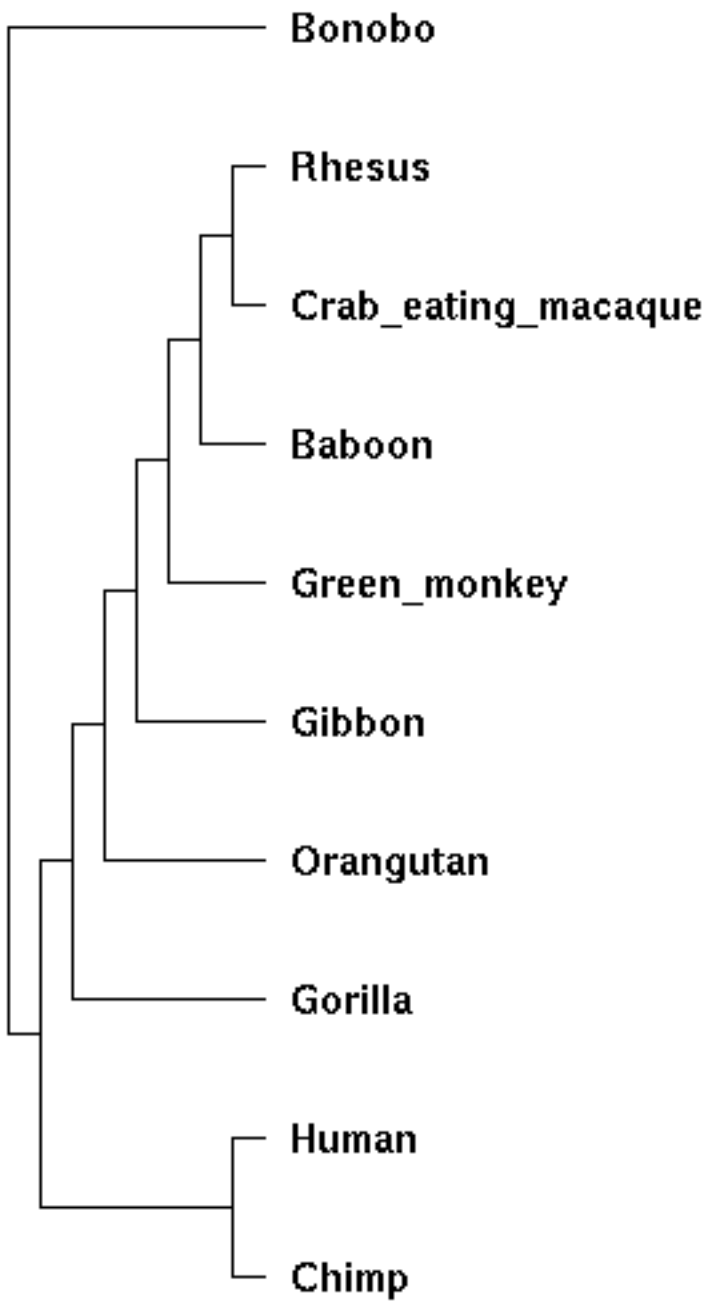
2a.

[1]

"(Bonobo:0.00392,(((((((Rhesus:0.004991,Crab_eating_macaque:0.004991):0.003,Baboon:0.00
8042):0.01961,Green_monkey:0.027):0.02204,Gibbon:0.02227):0.003471,Orangutan:0.01894):
0.009693,Gorilla:0.008964):0.000999999999999997,(Human:0.00655,Chimp:0.00684):0.00122
):0.00392)"

2b.

Use UCSC phyloPng tool:
Parameters:
Width:300
Height:512
Decimal places:5


Bonobo
Rhesus
Crab_eating_macaque
Baboon
Green_monkey
Gibbon
Orangutan
Gorilla
Human
Chimp