

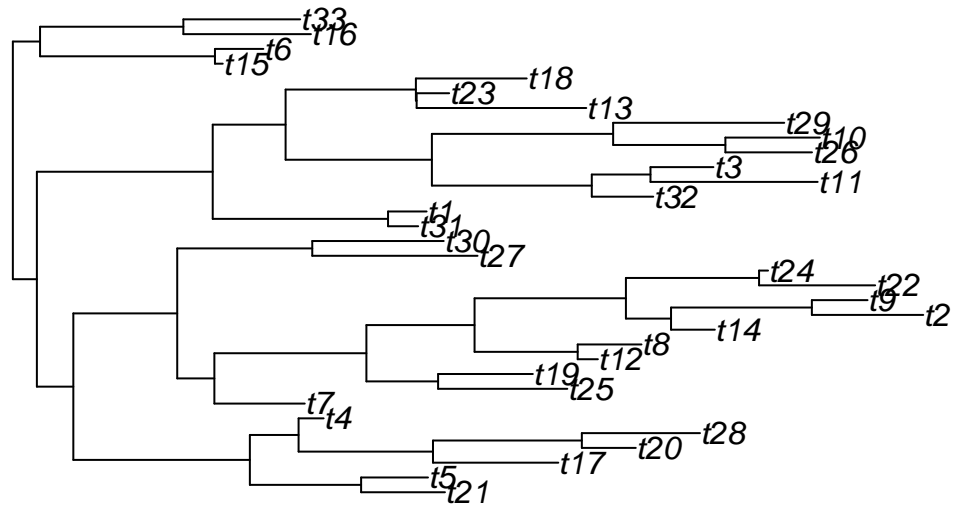
Lab2 732A51 Bioinformatics Group 9

Duc Duong, Martin Smelik, Raymond Sseguya

2018 M11 27

```
## 33 DNA sequences in binary format stored in a list.
##
## Mean sequence length: 1982.879
##   Shortest sequence: 931
##   Longest sequence: 2920
##
## Labels:
## JF806202
## HM161150
## FJ356743
## JF806205
## JQ073190
## GU457971
## ...
##
## Base composition:
##   a       c       g       t
## 0.312 0.205 0.231 0.252
## (Total: 65.44 kb)

## [1] "Created an AI gene and saved in file: AI_gene.fasta"
```



```
##          a          c          g          t
## a 0.2536379 0.2637915 0.2356308 0.2469398
## c 0.2643937 0.2373418 0.2233531 0.2749115
## g 0.2783559 0.2488010 0.2284553 0.2443878
## t 0.2576488 0.2355621 0.2672338 0.2395554
```

```
## For the sequences number: 1
```

```
## [1] "The composition of lizards dataset:"
```

```
##
```

```
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0501 0.0792 0.0000 0.0731 0.1032 0.2715 0.0000 0.1303 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0481 0.1553 0.0000 0.0862
```

```
## [1] "The composition of AI_gene1 dataset:"
```

```
##
```

```
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0982 0.2194 0.0000 0.0000 0.2184 0.4629 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
```

```
## [1] "The composition of Ai_gene2 dataset:"
```

```
##
```

```
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.060 0.066 0.069 0.064 0.067 0.063 0.048 0.064 0.063 0.064 0.064 0.052
##      ta      tc      tg      tt
## 0.069 0.050 0.062 0.074
```

```

##
## For the sequences number: 2
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.5814 0.0000 0.0000 0.1868 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.1868 0.0000 0.0000 0.0447
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9996 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.060 0.066 0.060 0.071 0.068 0.062 0.052 0.063 0.063 0.060 0.055 0.067
##      ta      tc      tg      tt
## 0.066 0.058 0.078 0.050
##
## For the sequences number: 3
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4825 0.2041 0.0000 0.0000 0.2041 0.1090 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4604 0.0000 0.0000 0.2196 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.2196 0.0000 0.0000 0.1000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.071 0.066 0.057 0.064 0.071 0.066 0.051 0.063 0.053 0.063 0.054 0.063
##      ta      tc      tg      tt
## 0.063 0.056 0.071 0.067
##
## For the sequences number: 4
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0525 0.0833 0.0714 0.0000 0.1021 0.2706 0.1298 0.0000 0.0525 0.1487
##      gg      gt      ta      tc      tg      tt
## 0.0823 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0466 0.1536 0.0000 0.0000 0.1536 0.6452

```

```

## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.078 0.088 0.047 0.060 0.068 0.079 0.061 0.064 0.058 0.048 0.053 0.058
##      ta      tc      tg      tt
## 0.069 0.057 0.057 0.054
##
## For the sequences number: 5
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0482 0.1879 0.0000 0.0000 0.1879 0.5753 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.5997 0.0000 0.0000 0.1716 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.1716 0.0000 0.0000 0.0563
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.075 0.066 0.053 0.078 0.070 0.075 0.052 0.056 0.062 0.056 0.053 0.056
##      ta      tc      tg      tt
## 0.065 0.056 0.069 0.057
##
## For the sequences number: 6
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0550 0.0000 0.0678 0.0843 0.0000 0.0000 0.0000 0.0000 0.0532 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0678 0.1366 0.0990 0.0000 0.1210 0.3144
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.9991 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.060 0.058 0.061 0.066 0.070 0.060 0.055 0.064 0.057 0.069 0.057 0.054
##      ta      tc      tg      tt
## 0.059 0.062 0.064 0.083
##
## For the sequences number: 7
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4815 0.2044 0.0000 0.0000 0.2044 0.1093 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

```

```

## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9997 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.057 0.058 0.061 0.058 0.071 0.073 0.064 0.077 0.057 0.074 0.054 0.054
##      ta      tc      tg      tt
## 0.049 0.081 0.059 0.052
##
## For the sequences number: 8
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.2679 0.2351 0.0000 0.0000 0.2341
##      gg      gt      ta      tc      tg      tt
## 0.2620 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.254 0.243
##      ta      tc      tg      tt
## 0.000 0.000 0.244 0.258
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.069 0.067 0.071 0.052 0.062 0.061 0.066 0.067 0.075 0.065 0.066 0.050
##      ta      tc      tg      tt
## 0.054 0.062 0.053 0.059
##
## For the sequences number: 9
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.4895 0.2107 0.0000 0.0000 0.2098 0.0891
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0881 0.2107 0.0000 0.0000 0.2107
##      gg      gt      ta      tc      tg      tt
## 0.4895 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.059 0.072 0.057 0.062 0.064 0.065 0.065 0.064 0.057 0.062 0.070 0.064
##      ta      tc      tg      tt
## 0.071 0.058 0.061 0.048
##
## For the sequences number: 10

```

```

## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4774 0.2065 0.0000 0.0000 0.2065 0.1092 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.1010 0.0000 0.0000 0.2116 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.2113 0.0000 0.0000 0.4757
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.076 0.072 0.061 0.058 0.079 0.053 0.060 0.059 0.057 0.067 0.056 0.062
##      ta      tc      tg      tt
## 0.055 0.059 0.065 0.060
##
## For the sequences number: 11
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0508 0.0653 0.0771 0.0110 0.0533 0.0717 0.0384 0.0877 0.0540 0.0632
##      gg      gt      ta      tc      tg      tt
## 0.1083 0.0863 0.0462 0.0508 0.0884 0.0472
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.1080 0.0824 0.0000 0.1378 0.0753 0.0597 0.0000 0.1087 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.1449 0.1012 0.0000 0.1818
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.053 0.064 0.059 0.070 0.065 0.077 0.067 0.049 0.067 0.065 0.063 0.063
##      ta      tc      tg      tt
## 0.060 0.052 0.069 0.056
##
## For the sequences number: 12
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4722 0.2059 0.0000 0.0000 0.2062 0.1092 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.9997
## [1] "The composition of Ai_gene2 dataset:"
##

```

```

##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.064 0.060 0.055 0.069 0.059 0.042 0.068 0.051 0.066 0.066 0.065 0.067
##      ta      tc      tg      tt
## 0.059 0.052 0.076 0.080
##
## For the sequences number: 13
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.6426 0.0000 0.1536 0.0000 0.0000 0.0000 0.0000 0.0000 0.1536 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0499 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9996 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.065 0.063 0.072 0.053 0.066 0.057 0.063 0.051 0.060 0.070 0.073 0.071
##      ta      tc      tg      tt
## 0.062 0.048 0.065 0.060
##
## For the sequences number: 14
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4780 0.2062 0.0000 0.0000 0.2062 0.1079 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9996 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.074 0.061 0.057 0.070 0.058 0.048 0.054 0.065 0.062 0.054 0.064 0.063
##      ta      tc      tg      tt
## 0.068 0.061 0.068 0.072
##
## For the sequences number: 15
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.6419 0.0000 0.1521 0.0000 0.0000 0.0000 0.0000 0.0000 0.1517 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0509 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##

```

```

##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.6416 0.0000 0.1590 0.0000 0.0000 0.0000 0.0000 0.0000 0.1587 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0404 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.065 0.066 0.052 0.062 0.057 0.070 0.058 0.062 0.063 0.054 0.067 0.059
##      ta      tc      tg      tt
## 0.060 0.057 0.066 0.081
##
## For the sequences number: 16
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.2709 0.0956 0.1365 0.0000 0.1454
##      gg      gt      ta      tc      tg      tt
## 0.0418 0.0498 0.0000 0.0857 0.1006 0.0707
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.6016 0.1683 0.0000 0.0000 0.1693 0.0598
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.068 0.070 0.057 0.071 0.050 0.058 0.052 0.077 0.066 0.055 0.042 0.059
##      ta      tc      tg      tt
## 0.081 0.055 0.071 0.067
##
## For the sequences number: 17
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.053 0.081 0.000 0.071 0.095 0.272 0.000 0.135 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.056 0.150 0.000 0.084
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.000 0.000 0.000 0.000 0.000 0.053 0.000 0.162 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.000 0.162 0.000 0.622
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.056 0.063 0.081 0.047 0.059 0.067 0.066 0.068 0.067 0.069 0.057 0.064
##      ta      tc      tg      tt
## 0.065 0.061 0.053 0.056
##
## For the sequences number: 18
## [1] "The composition of lizards dataset:"
##

```



```

##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.2411 0.1179 0.0935 0.0000 0.1165 0.1112 0.0927 0.0000 0.0949 0.0913
##      gg      gt      ta      tc      tg      tt
## 0.0407 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.2959 0.0000 0.2552 0.0000 0.0000 0.0000 0.0000 0.0000 0.2552 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.1933 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.053 0.064 0.056 0.068 0.068 0.070 0.062 0.071 0.064 0.070 0.049 0.053
##      ta      tc      tg      tt
## 0.056 0.066 0.069 0.060
##
## For the sequences number: 19
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4770 0.2043 0.0000 0.0000 0.2036 0.1120 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4514 0.0000 0.0000 0.2257 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.2257 0.0000 0.0000 0.0969
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.069 0.060 0.059 0.070 0.068 0.052 0.065 0.051 0.064 0.064 0.062 0.065
##      ta      tc      tg      tt
## 0.057 0.061 0.069 0.063
##
## For the sequences number: 20
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0536 0.0784 0.0724 0.0000 0.0982 0.2679 0.1349 0.0000 0.0526 0.1538
##      gg      gt      ta      tc      tg      tt
## 0.0873 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0317 0.1607 0.0000 0.0000 0.1607 0.6458
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.075 0.063 0.071 0.066 0.079 0.062 0.049 0.056 0.054 0.064 0.057 0.062

```

```

##      ta      tc      tg      tt
## 0.067 0.058 0.059 0.057
##
## For the sequences number: 21
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.9993 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.9993
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.056 0.050 0.065 0.057 0.064 0.067 0.058 0.067 0.048 0.071 0.064 0.073
##      ta      tc      tg      tt
## 0.059 0.068 0.069 0.063
##
## For the sequences number: 22
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4691 0.2084 0.0000 0.0000 0.2081 0.1134 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9997 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.066 0.070 0.052 0.058 0.056 0.079 0.070 0.062 0.065 0.057 0.061 0.057
##      ta      tc      tg      tt
## 0.060 0.060 0.057 0.069
##
## For the sequences number: 23
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0435 0.1811 0.0000 0.0000 0.1811 0.5895 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0449 0.0000 0.0000 0.1686 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

```

```

##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.1679 0.0000 0.0000 0.6178
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.065 0.055 0.067 0.058 0.058 0.067 0.056 0.060 0.061 0.058 0.063 0.075
##      ta      tc      tg      tt
## 0.062 0.061 0.071 0.062
##
## For the sequences number: 24
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.053 0.082 0.000 0.079 0.107 0.253 0.000 0.129 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.054 0.154 0.000 0.088
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.000 0.000 0.000 0.000 0.000 0.241 0.000 0.247 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.000 0.247 0.000 0.264
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.047 0.063 0.067 0.047 0.065 0.058 0.073 0.053 0.063 0.071 0.083 0.077
##      ta      tc      tg      tt
## 0.049 0.057 0.071 0.055
##
## For the sequences number: 25
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4748 0.2073 0.0000 0.0000 0.2073 0.1103 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9996 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.069 0.057 0.062 0.054 0.066 0.060 0.063 0.060 0.058 0.067 0.065 0.073
##      ta      tc      tg      tt
## 0.049 0.065 0.072 0.059
##
## For the sequences number: 26
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.6301 0.0000 0.1584 0.0000 0.0000 0.0000 0.0000 0.0000 0.1584 0.0000

```

```

##      gg      gt      ta      tc      tg      tt
## 0.0527 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.9996 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.067 0.065 0.055 0.056 0.059 0.065 0.066 0.061 0.070 0.064 0.054 0.060
##      ta      tc      tg      tt
## 0.047 0.057 0.073 0.080
##
## For the sequences number: 27
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.5453 0.0000 0.1910 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.1910 0.0000 0.0716
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.000 0.000 0.000 0.999
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.044 0.069 0.051 0.065 0.056 0.048 0.072 0.064 0.064 0.068 0.079 0.061
##      ta      tc      tg      tt
## 0.065 0.055 0.070 0.068
##
## For the sequences number: 28
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4767 0.2044 0.0000 0.0000 0.2048 0.1103 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4784 0.0000 0.0000 0.2100 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.2100 0.0000 0.0000 0.1013
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.065 0.074 0.046 0.052 0.058 0.076 0.065 0.072 0.064 0.056 0.057 0.062
##      ta      tc      tg      tt
## 0.051 0.065 0.070 0.066

```

```

##
## For the sequences number: 29
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4725 0.2058 0.0000 0.0000 0.2058 0.1148 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.4659 0.0000 0.0000 0.2152 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.2148 0.0000 0.0000 0.1038
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.065 0.060 0.049 0.064 0.052 0.065 0.066 0.068 0.060 0.056 0.074 0.059
##      ta      tc      tg      tt
## 0.061 0.070 0.060 0.070
##
## For the sequences number: 30
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.2406 0.2274 0.0000 0.0000 0.2264 0.3047
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.9991 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.072 0.064 0.054 0.066 0.050 0.081 0.068 0.065 0.068 0.063 0.049 0.053
##      ta      tc      tg      tt
## 0.066 0.055 0.062 0.063
##
## For the sequences number: 31
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0698 0.1278 0.0000 0.0511 0.1129 0.3265 0.0000 0.1037 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0668 0.0888 0.0000 0.0515
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.2235 0.0000 0.2396 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0000 0.2392 0.0000 0.2973

```

```

## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.064 0.073 0.051 0.059 0.066 0.081 0.067 0.060 0.060 0.067 0.054 0.056
##      ta      tc      tg      tt
## 0.057 0.053 0.064 0.067
##
## For the sequences number: 32
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0489 0.0818 0.0000 0.0728 0.0977 0.2702 0.0000 0.1326 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.0000 0.0000 0.0568 0.1476 0.0000 0.0907
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##      ta      tc      tg      tt
## 0.000 0.000 0.000 0.999
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.059 0.062 0.055 0.059 0.066 0.087 0.066 0.062 0.060 0.059 0.052 0.061
##      ta      tc      tg      tt
## 0.050 0.073 0.060 0.068
##
## For the sequences number: 33
## [1] "The composition of lizards dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0537 0.0730 0.0763 0.0000 0.0945 0.2760 0.1332 0.0000 0.0548 0.1536
##      gg      gt      ta      tc      tg      tt
## 0.0838 0.0000 0.0000 0.0000 0.0000 0.0000
## [1] "The composition of AI_gene1 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc
## 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
##      gg      gt      ta      tc      tg      tt
## 0.2470 0.2342 0.0000 0.0000 0.2342 0.2836
## [1] "The composition of Ai_gene2 dataset:"
##
##      aa      ac      ag      at      ca      cc      cg      ct      ga      gc      gg      gt
## 0.063 0.058 0.080 0.057 0.058 0.054 0.068 0.055 0.084 0.054 0.074 0.064
##      ta      tc      tg      tt
## 0.053 0.069 0.055 0.053

## $estimate
## MLE Fit
## A 4 - dimensional discrete Markov Chain defined by the following states:
## a, c, g, t
## The transition matrix (by rows) is defined as follows:
##      a      c      g      t
## a 0.6671729 0.2394909 0.06098894 0.03234722

```

```

## c 0.3756621 0.5070469 0.05169315 0.06559781
## g 0.2987863 0.2020121 0.35499840 0.14420313
## t 0.2284000 0.2542000 0.18820000 0.32920000
##
##
## $standardError
##           a           c           g           t
## a 0.004501558 0.002697045 0.001361035 0.0009912021
## c 0.004215073 0.004897004 0.001563590 0.0017613720
## g 0.006907547 0.005679789 0.007529336 0.0047987797
## t 0.006758698 0.007130217 0.006135145 0.0081141851
##
## $confidenceLevel
## [1] 0.95
##
## $lowerEndpointMatrix
##           a           c           g           t
## a 0.6597685 0.2350547 0.05875024 0.03071684
## c 0.3687289 0.4989921 0.04912127 0.06270061
## g 0.2874244 0.1926697 0.34261375 0.13630984
## t 0.2172829 0.2424718 0.17810859 0.31585335
##
## $upperEndpointMatrix
##           a           c           g           t
## a 0.6745773 0.2439272 0.06322765 0.03397761
## c 0.3825953 0.5151018 0.05426503 0.06849500
## g 0.3101482 0.2113546 0.36738306 0.15209642
## t 0.2395171 0.2659282 0.19829141 0.34254665

## $estimate
## MLE Fit
## A 4 - dimensional discrete Markov Chain defined by the following states:
## a, c, g, t
## The transition matrix (by rows) is defined as follows:
##           a           c           g           t
## a 0.8377184 0.01302527 0.03344404 0.1158123
## c 0.1178405 0.44614963 0.06029049 0.3757194
## g 0.1698972 0.03230543 0.65859031 0.1392070
## t 0.1981986 0.06639433 0.04675657 0.6886505
##
##
## $standardError
##           a           c           g           t
## a 0.004918743 0.000613336 0.0009827985 0.001828868
## c 0.005682774 0.011057408 0.0040647840 0.010147168
## g 0.004994817 0.002178032 0.0098340885 0.004521235
## t 0.003123272 0.001807694 0.0015169835 0.005821822
##
## $confidenceLevel
## [1] 0.95
##
## $lowerEndpointMatrix
##           a           c           g           t
## a 0.8296278 0.01201642 0.03182748 0.1128041

```

```

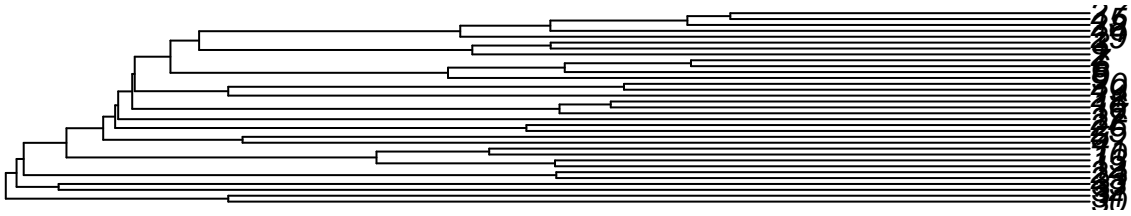
## c 0.1084932 0.42796181 0.05360452 0.3590288
## g 0.1616815 0.02872289 0.64241467 0.1317703
## t 0.1930613 0.06342094 0.04426135 0.6790744
##
## $upperEndpointMatrix
##      a      c      g      t
## a 0.8458090 0.01403412 0.03506060 0.1188205
## c 0.1271878 0.46433745 0.06697647 0.3924100
## g 0.1781130 0.03588798 0.67476594 0.1466438
## t 0.2033360 0.06936772 0.04925179 0.6982265

## $estimate
## MLE Fit
## A 4 - dimensional discrete Markov Chain defined by the following states:
## a, c, g, t
## The transition matrix (by rows) is defined as follows:
##      a      c      g      t
## a 0.2558535 0.2587650 0.2389907 0.2463909
## c 0.2513208 0.2579251 0.2444765 0.2462776
## g 0.2519261 0.2531491 0.2458114 0.2491134
## t 0.2413919 0.2406619 0.2625624 0.2553839
##
##
## $standardError
##      a      c      g      t
## a 0.005571254 0.005602865 0.005384530 0.005467259
## c 0.005493434 0.005565144 0.005418114 0.005438036
## g 0.005550596 0.005564052 0.005482821 0.005519523
## t 0.005419408 0.005411207 0.005652060 0.005574260
##
## $confidenceLevel
## [1] 0.95
##
## $lowerEndpointMatrix
##      a      c      g      t
## a 0.2466896 0.2495491 0.2301339 0.2373980
## c 0.2422850 0.2487712 0.2355645 0.2373328
## g 0.2427962 0.2439970 0.2367930 0.2400346
## t 0.2324778 0.2317612 0.2532655 0.2462150
##
## $upperEndpointMatrix
##      a      c      g      t
## a 0.2650173 0.2679809 0.2478474 0.2553837
## c 0.2603567 0.2670789 0.2533885 0.2552224
## g 0.2610561 0.2623011 0.2548299 0.2581922
## t 0.2503060 0.2495625 0.2718592 0.2645527

## $lambda
## [1] 0.5 0.5
##
## $Q
## $Q[[1]]
##      a      c      t <NA>
## a 0.2475248 0.203960396 0.1660900 1

```


UPGMA



NJ

