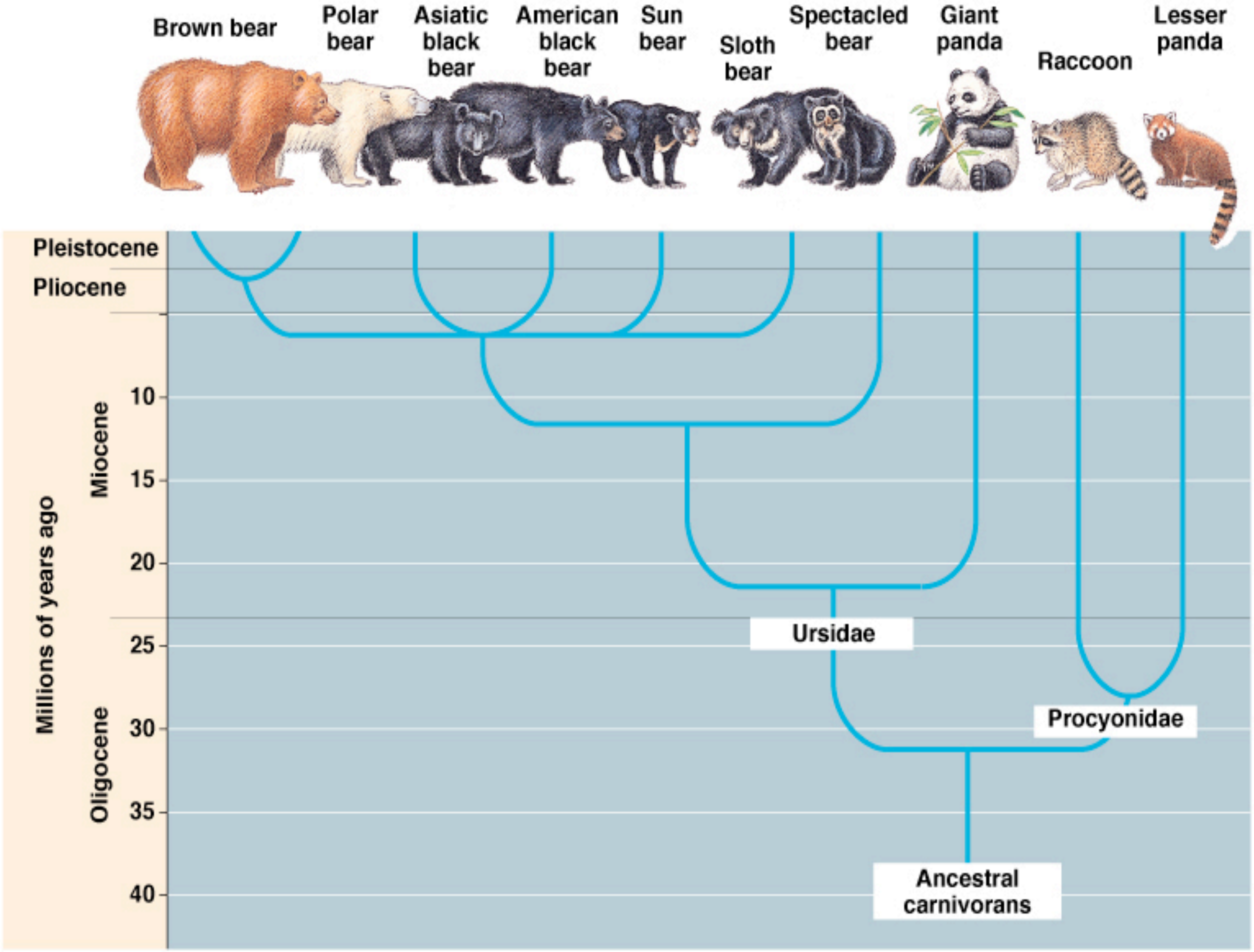
Play with Phylogenetic Trees

Ge Gao and Xiying Wang
Center of Bioinformatics, PKU

2005-12-31

Outline

- What's Phylogenetic Trees?
- Build Phylogenetic Trees by Distance Methods
- Validate Phylogenetic Trees by Re-sampling
- Rock with PHYLIP


Brown bear
Polar bear
Asiatic black bear
American black bear
Sun bear
Sloth bear
Spectacled bear
Giant panda
Raccoon
Lesser panda
Pleistocene
Pliocene
Miocene
Oligocene
Millions of years ago
10
15
20
25
30
35
40
Ursidae
Procyonidae
Ancestral carnivorans

Phylogenetic Trees

- *Phylogenetics* is the study of evolutionary relationships among organisms

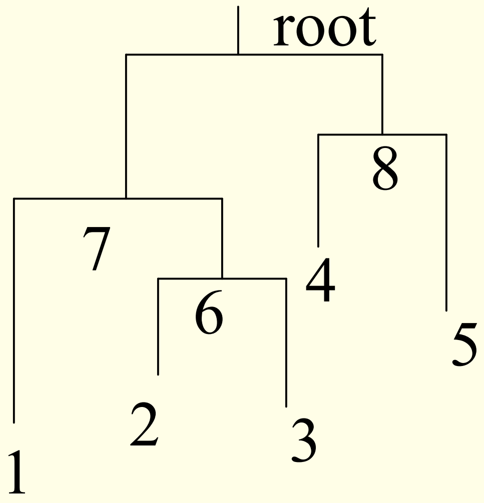
- A *phylogenetic tree* or *phylogeny* for a set of taxa (species, genes, …) is an evolutionary tree representing their relationships.
  - A tree is an acyclic graph: horizontal transfer is ignored
  - Edge weights *may* represent distance in evolution

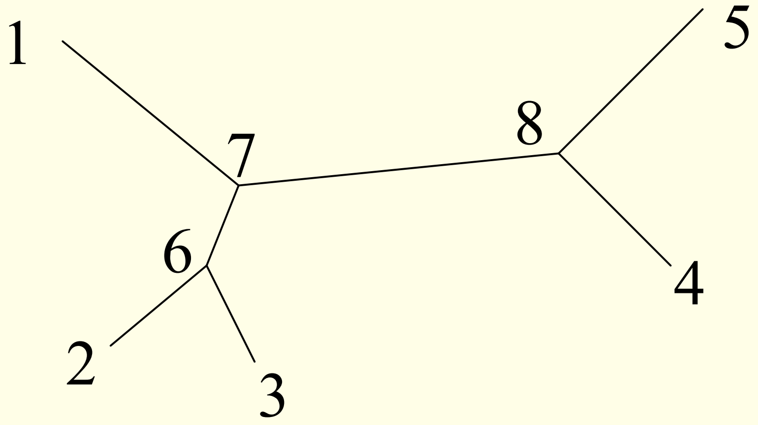
Phylogenetic Trees

- Trees can be rooted or unrooted.
  - In the case of unrooted trees we can assume to have not enough data to determine the root of the tree

- The leaves of a phylogenetic tree usually represent the present day taxa, the internal nodes represent hypothesized ancestors.

Tree Topology





Why Phylogenetic Trees?

- Evolution of organisms （tree of species)
- Evolution of genes (tree of gene)
- Application:
  - Comparative Genomics
  - Gene function prediction

Models and Methods

- Model: an abstract of "*real*" evolutionary events.

- Maximum Parsimony methods
- Distance Matrix methods
- Maximum Likelihood methods

- Which is better?

Maximum Parsimony

- Variation is small

- All possible trees are evaluated
  - <=11 or 12 sequences concerned
  - Time-consuming

- Concensus tree for more than one MP trees

Distance Matrix methods

- Variation is intermediate

- Hierarchical inference
  - Rather faster then MP.
  - Large number of sequences

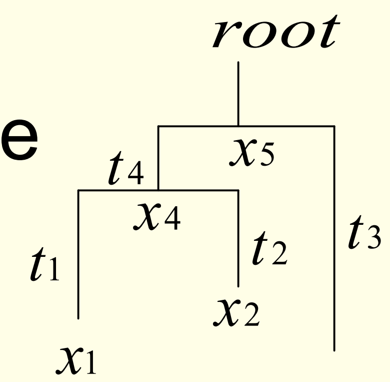
- The distance matrix can be derived from multiple alignment or evolution event or others like K-tuple method

Maximum Likelihood

- Variation could be some larger

- All possible trees are evaluated
  - <=11 or 12 sequences concerned

- Both topology and edge lengths are considered.
  - based on probability inference.



$$P(x^{\bullet} \mid T, t_{\bullet})$$

How many possible trees?

Rooted tree

$$\frac{(2m-3)!}{2^{m-2}\cdot(m-2)!}$$

m=10: 34,459,425

Unrooted tree

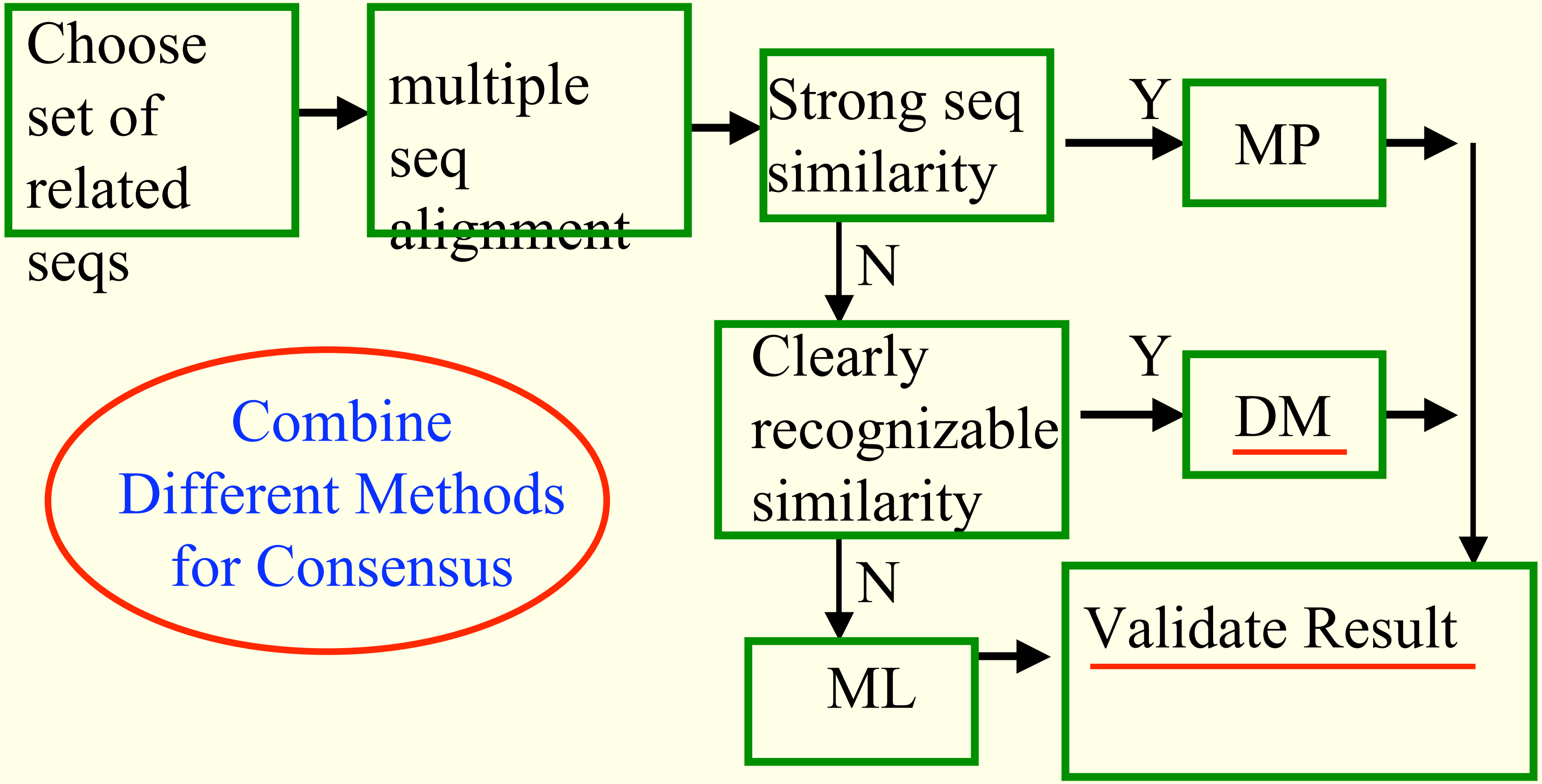
$$\frac{(2m-5)!}{2^{m-3}\cdot(m-3)!}$$

m=10: 2,027,025

A Quick Summary

	MP	DM	ML
Variation	+	++	+++
Computation Complex	++	+	+++
Edge Length Estimation	N	N	Y
Flexibility	+	+++	++

A General Protocol


Choose set of related seqs
multiple seq alignment
Strong seq similarity
Y
MP
N
Clearly recognizable similarity
Y
DM
N
ML
Validate Result
Combine Different Methods for Consensus

Outline

- What’s Phylogenetic Trees?
- Build Phylogenetic Trees by Distance Methods
- Validate Phylogenetic Trees by Re-sampling
- Rock by PHYLIP

Distance Methods

- Neighbors – the closest taxa
- Rather fast
- More reliable than MP when branch lengths vary (Jin and Nei, 1990; Swofford et al. 1996)
- Additive: the lengths be additive

Neighbors Joining

- Proposed by Saitou and Nei in 1987
  - Pearson et al. enhance NJ in 1999 (Not a single tree predicted)
- Pairing sequences based on the effect of the pairing on the sum of the sum of the branch lengths of the tree
- Starting from a star-like tree

Similarity to Distance

- Convert alignment scores to distances:

$$D = -\log S_{eff} = -\log\{(S_{obs} - S_{rand})/(S_{\max} - S_{rand})\}$$

S_{obs} is observed pairwise alignment score

$S_{\max}$ is the maximum score, the average of the score of aligning either sequence to itself.

$\boxed{S_{rand}}$ is the expected score for aligning two random sequences of the same length and residue composition, which can be calculated by random shuffling of the two sequences or by an approximate calculation given in Feng & Doolittle[1996]

Neighbour Joining Algorithm

- For each node i the distance from the rest of the tree is estimated by

$$\boxed{r_i = \frac{1}{N-2} \sum_{k \neq i} d_{i,k}}$$

- Choose the nodes *i* and *j* that for which

$D_{ij} = d_{ij} - r_i - r_j$ is smallest

join i and j (ij is new node)

- Compute branch length from i and j to ij

$$d_{i,(ij)} = \frac{1}{2} d_{i,j} + \frac{1}{2}(r_i - r_j), d_{j,(ij)} = \frac{1}{2} d_{i,j} + \frac{1}{2}(r_j - r_i)$$

- Compute the distances between the new cluster and each other cluster:

$$d_{(ij),k} = \frac{d_{i,k} + d_{j,k} - d_{i,j}}{2}$$

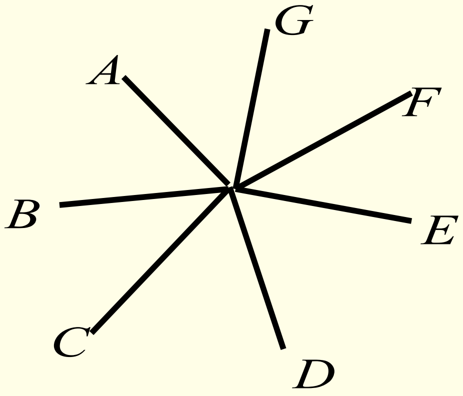
Neighbour joining algorithm(1)

	A	B	C	D	E	F	G	r_i
A		63	94	111	67	23	107	88.4
B	63		79	96	16	58	92	80.8
C	94	79		47	83	89	43	87
D	111	96	47		100	106	20	96
E	67	16	83	100		62	96	84.4
F	23	58	89	106	62		102	88
G	107	92	43	20	96	102		92

No molecular clock assumption

Start from the star-like tree

Calculate r_i



Neighbour joining algorithm(2)

	A	B	C	D	E	F	G	r_i
A		-106.2	-81.4	-73.4	-105.8	-153.4	-69.4	88.4
B	63		-88.8	-80.8	-149.2	-110.8	-80.8	80.8
C	94	79		-136	-84.4	-86	-136	87
D	111	96	47		-80.4	-78	-168	96
E	67	16	83	100		-110.4	-80.4	84.4
F	23	58	89	106	62		-78	88
G	107	92	43	20	96	102		92

Calculate D_{ij} , D and G are the closest

Calculate the branch lengths of D and G

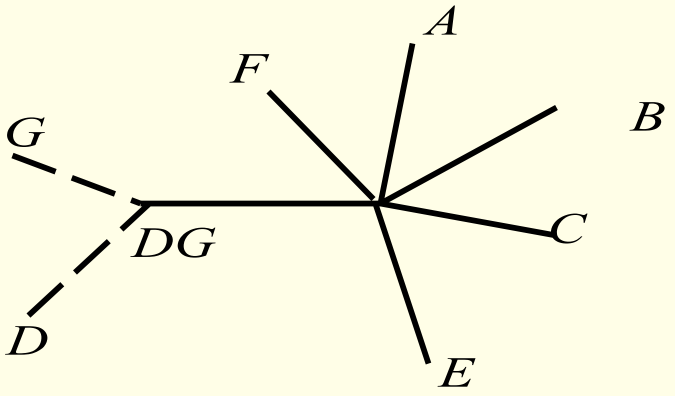
$d = 12$

$g = 8$

Neighbour joining algorithm(3)

	A	B	C	E	F	DG	r_i
A		63	94	67	23	94	85.25
B	63		79	16	58	84	75
C	94	79		83	89	35	95
E	67	16	83		62	88	79
F	23	58	89	62		94	81.5
DG	94	84	35	88	94		91.25

Join D and G, calculate the distances r_i from DG to other nodes



Neighbour joining algorithm(4)

	A	B	C	E	F	DG	r_i
A		-97.25	-86.25	-97.25	-143.75	-82.5	85.25
B	63		-91	-138	-98.5	-82.25	75
C	94	79		-91	-87.5	-151.25	95
E	67	16	83		-98.5	-82.25	79
F	23	58	89	62		-78.75	81.5
DG	94	84	35	88	94		91.25

Calculate D_{ij} , C and DG are the closest

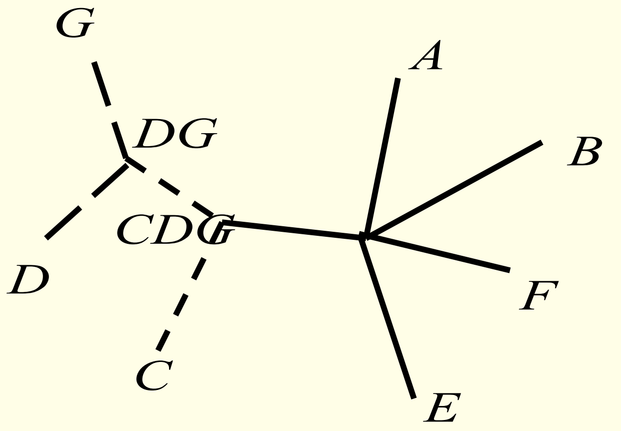
Calculate the branch lengths of C and DG

$$c = 19.375$$

$$dg = 15.625$$

Neighbour joining algorithm(5)

	A	B	E	F	CDG	r_i
A		63	67	23	61	71.3
B	63		16	58	64	67
E	67	16		62	60	68.3
F	23	58	62		74	72.3
CDG	61	64	60	74		98.3



Join DG and C, calculate the distances r_i from CDG to other nodes

Neighbour joining algorithm(6)

	A	B	E	F	CDG	r_i
A		-75.3	-72.6	-120.6	-108.6	71.3
B	63		-119.3	-81.3	-101.3	67
E	67	16		-78.6	-90	68.3
F	23	58	62		-96.3	72.3
CDG	61	64	60	74		98.3

Calculate D_{ij} , A and F are the closest

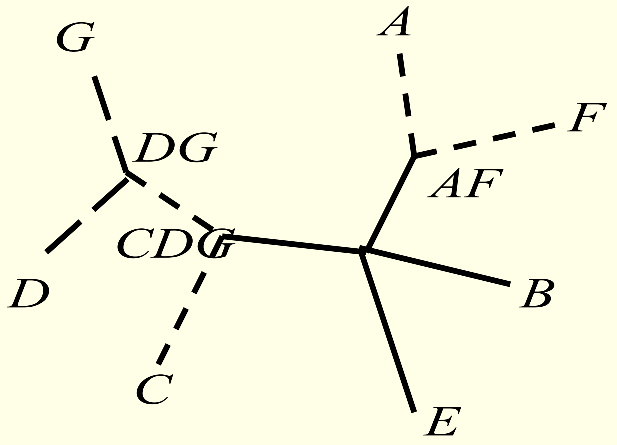
Calculate the branch lengths of A and F

$$a = 11$$

$$f = 12$$

Neighbour joining algorithm(7)

	AF	B	E	CDG	r_i
AF		98	106	112	158
B	98		16	64	89
E	106	16		60	91
CDG	112	64	60		118



Join A and F, calculate the distances r_i from AF to other nodes

Neighbour joining algorithm(8)

	AF	B	E	CDG	r_i
AF		-149	-143	-164	158
B	98		-164	-143	89
E	106	16		-149	91
CDG	112	64	60		118

Calculate D_{ij} , B and E are the closest

Calculate the branch lengths of B and E

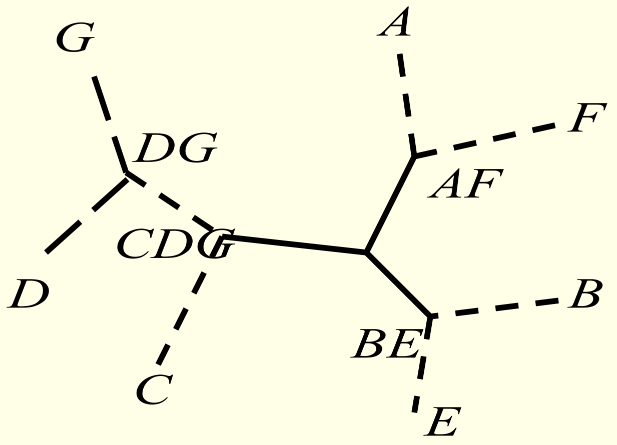
$$b = 7$$

$$e = 9$$

Neighbour joining algorithm(9)

	AF	BE	CDG	r_i
AF		188	112	300
BE	188		108	296
CDG	112	108		220

Join B and E, calculate the distances from BE to other nodes and r_i



Neighbour joining algorithm(10)

	AF	BE	CDG	r_i
AF		-408	-408	300
BE	188		-408	296
CDG	112	108		220

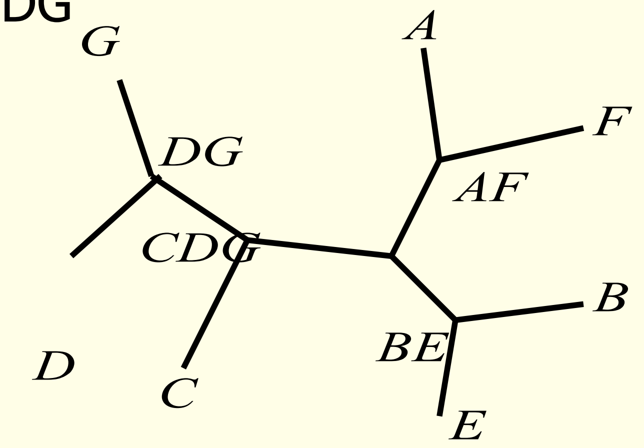
Calculate D_{ij}, BE and CDG are the closest

Calculate the branch lengths of BE and CDG

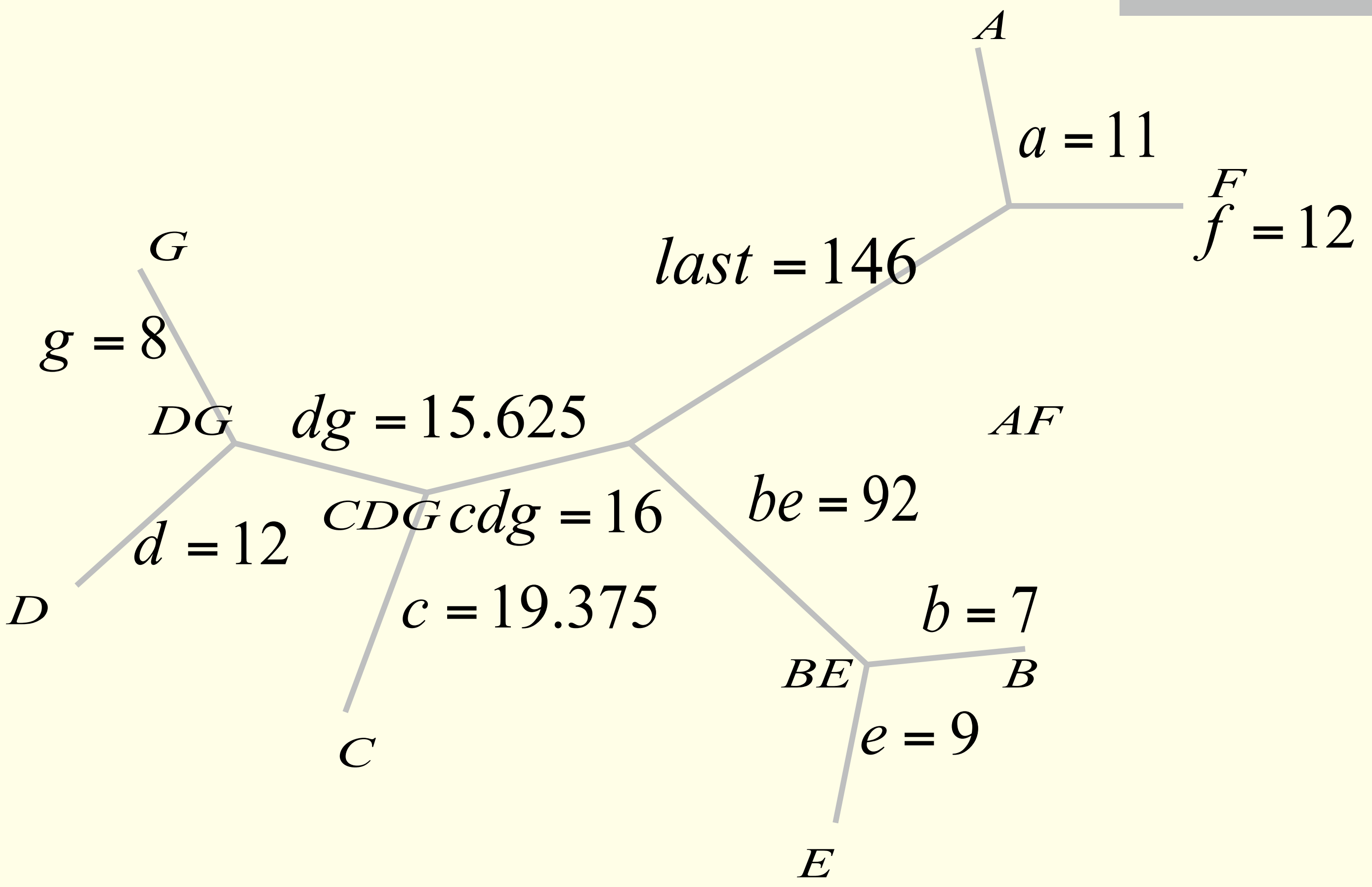
$be = 92$

$cdg = 16$

Join BE and CDG, calculate the distances from BECDG to the last node AF :146



Neighbour joining algorithm(11)

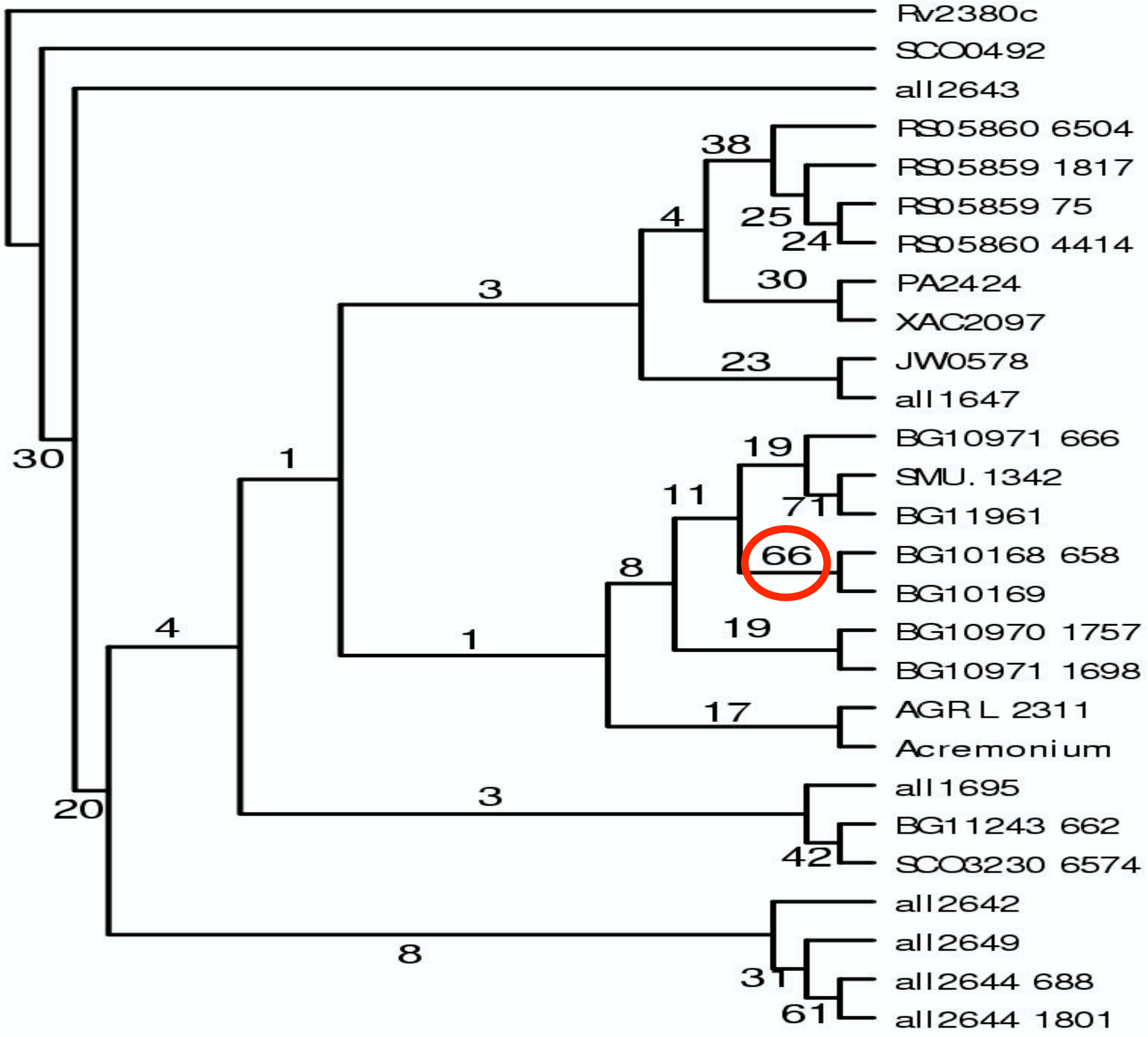


A Quick Summary

- NJ is fast and reliable for topology
  - But not edges length
- NJ do not necessarily assume molecular clock.
  - But it guarantees the assumption hold if required.
- Distances should hold Triangle Law.

Outline

- What’s Phylogenetic Trees?
- Build Phylogenetic Trees by Distance Methods
- Validate Phylogenetic Trees by Re-sampling
- Rock with PHYLIP


Rv2380c
SCO0492
all2643
RS05860 6504
RS05859 1817
RS05859 75
RS05860 4414
PA2424
XAC2097
JW0578
all1647
BG10971 666
SMU.1342
BG11961
BG10168 658
BG10169
BG10970 1757
BG10971 1698
AGR L 2311
Acremonium
all1695
BG11243 662
SCO3230 6574
all2642
all2649
all2644 688
all2644 1801
38
25
24
4
30
3
23
19
71
11
66
8
19
1
1
17
30
4
3
42
20
8
31
61

Validate the Inference

- Phylogenetic trees are inferred based on Model
  - Hypothetical Inference

- How reliable are the result?
  - Reliability vs. Stability
  - Validate the result by Re-sampling.

Bootstrap(1)

- Given a dataset consisting of an alignment of sequences, an artificial dataset of the same size is generated
  - by picking columns from the alignment at random with replacement.
- One given column in the original dataset can therefore appear several times in the artificial dataset

Bootstrap(2)

- The tree building algorithm is then applied to this new dataset, and the whole selection and tree building procedure is repeated typically 100 times.
- The frequency with which a chosen phylogenetic feature appears is taken to be a measure of the confidence we can have in this feature.
- At last, a consensus tree is created

Validate the Tree

- To improve prediction of trees and assist with localization of the root, an outgroup could be set.

- An outgroup of the following criteria:
  - From species that are known to have separated from the others at an early evolutionary time
  - More distantly related with other sequences

More words on Outgroup

- More than one can be selected
- By independently information, such as fossil evidence
- Too distant an outgroup may lead to incorrect prediction

Outline

- What’s Phylogenetic Trees?
- Build Phylogenetic Trees by Distance Methods
- Validate Phylogenetic Trees by Re-sampling
- Rock with PHYLIP

Phylogenetic Software

- Multialignment
  - ClustalW
  - POA

- Phylogenetic analysis
  - PHYLIP (Felsenstein,1989,1996)
  - PAUP (Sinauar Associates)
  - PAML (Yang Ziheng)
  - MEGA (Nei)
  - MacClade (Macintosh computer)

Programs in PHYLIP

- Create a distance table by:
  - DNADIST: various models of evolution
  - PROTDIST: based on the PAM model or others

- as input to the following:
  - NEIGHBOR:
    - NJ, no clock, no root
    - UPGMA and a clock and root

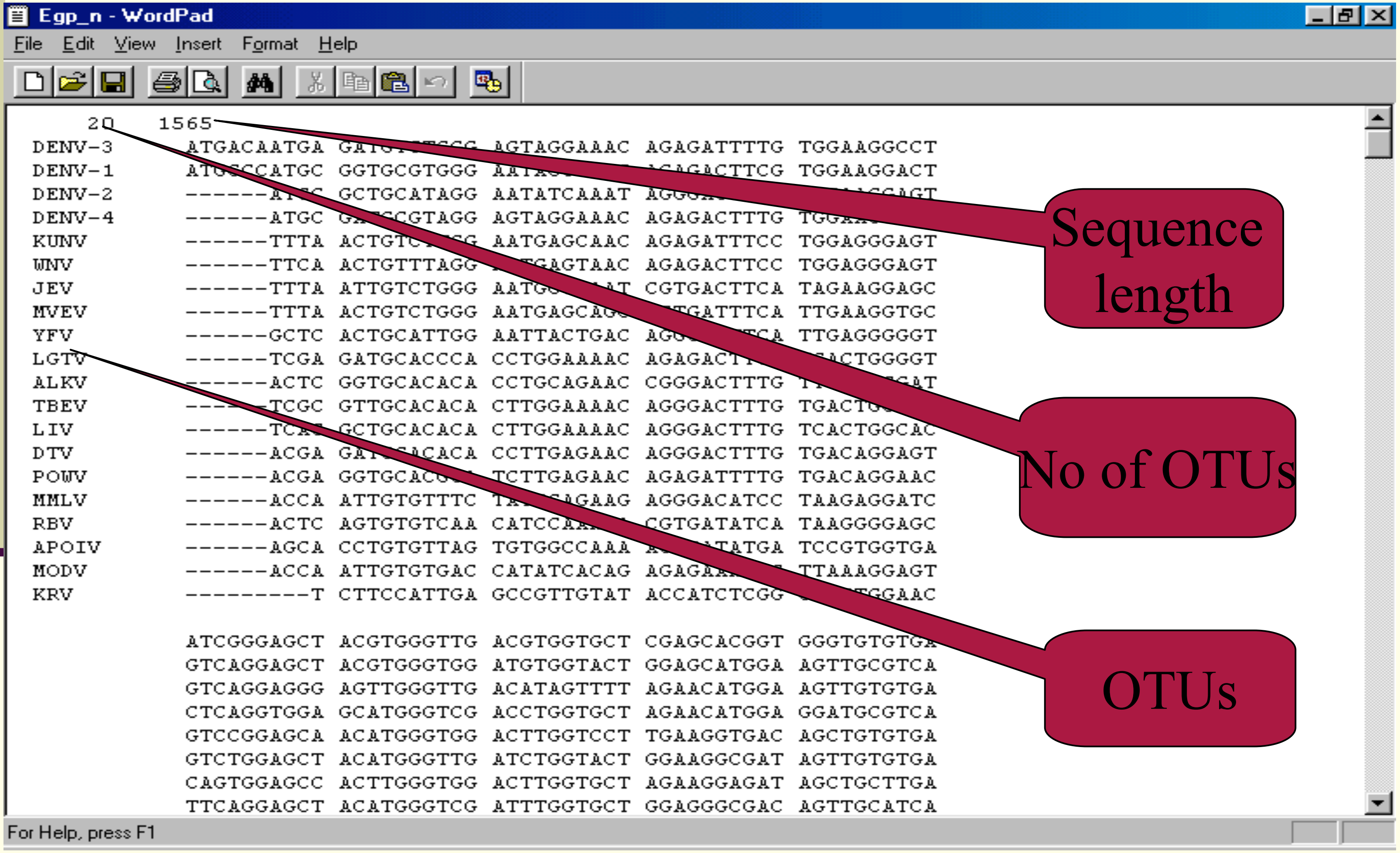
NJ @ PHYLIP

- Multiple alignment: clustalw,
  - save the output in phylip format (*.phy)
- Bootstrap the sequence data: SEQBOOT
- Build Phylogenetic trees: NEIGHBOR
- Calc Consensus : CONSENSUS

- Mo3 ATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGCACGGTACCAT
- Mo5 ATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCAT
- Mo6 ATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCAT
- Mo7 ATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCAT
- Mo8 ATGTATTTCGTACATTACTGCCAGCCACCATGAATATTGTACAGTACCAT
- Mo9 ATGTATCTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCAT
- Mo12 ATGTATTTCGTACATTACTG CCAGCCACCATGAATATTGTACGGTACCAT
- Mo13 ATGTATCTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCAT

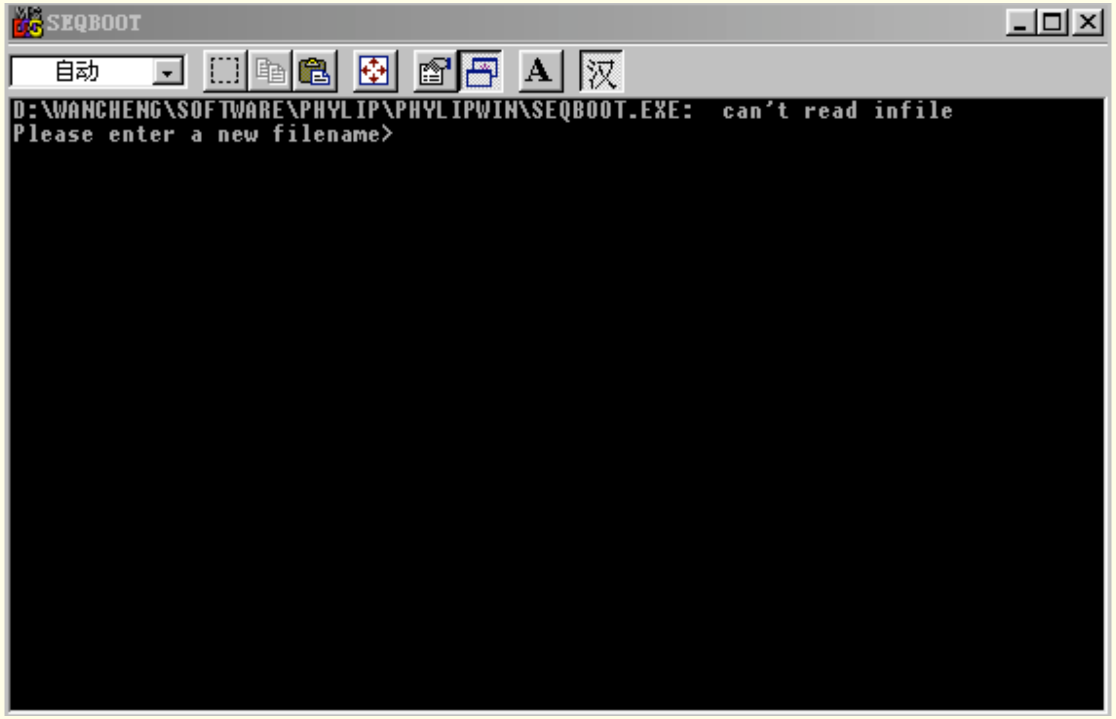
Mutiple Sequence Alignment (*.PHY)

Multiple alignment in Phylip format



```
    20    1565
 DENV-3     ATGACAATGA GAT[illegible]CGG AGTAGGAAAC AGAGATTTTG TGGAAGGCCT
 DENV-1     ATG[illegible]CCATGC GGTGCGTGGG AATA[illegible] [illegible]AGACTTCG TGGAAGGACT
 DENV-2     ------AT[illegible]C GCTGCATAGG AATATCAAAT AGG[illegible] [illegible]AAGGAGT
 DENV-4     ------ATGC GA[illegible]CGTAGG AGTAGGAAAC AGAGACTTTG TGGA[illegible]
 KUNV       ------TTTA ACTGTC[illegible]GG AATGAGCAAC AGAGATTTCC TGGAGGGAGT
 WNV        ------TTCA ACTGTTTAGG [illegible]TGAGTAAC AGAGACTTCC TGGAGGGAGT
 JEV        ------TTTA ATTGTCTGGG AATG[illegible]AT CGTGACTTCA TAGAAGGAGC
 MVEV       ------TTTA ACTGTCTGGG AATGAGCAG[illegible] [illegible]TGATTTCA TTGAAGGTGC
 YFV        ------GCTC ACTGCATTGG AATTACTGAC AG[illegible]TCA TTGAGGGGGT
 LGTV       ------TCGA GATGCACCCA CCTGGAAAAC AGAGACT[illegible] [illegible]ACTGGGGT
 ALKV       ------ACTC GGTGCACACA CCTGCAGAAC CGGGACTTTG T[illegible]GAT
 TBEV       ------TCGC GTTGCACACA CTTGGAAAAC AGGGACTTTG TGACTG[illegible]
 LIV        ------TCA[illegible] GCTGCACACA CTTGGAAAAC AGGGACTTTG TCACTGGCAC
 DTV        ------ACGA GA[illegible]ACACA CCTTGAGAAC AGGGACTTTG TGACAGGAGT
 POWV       ------ACGA GGTGCACG[illegible] TCTTGAGAAC AGAGATTTTG TGACAGGAAC
 MMLV       ------ACCA ATTGTGTTTC TA[illegible]AGAAG AGGGACATCC TAAGAGGATC
 RBV        ------ACTC AGTGTGTCAA CATCCA[illegible] CGTGATATCA TAAGGGGAGC
 APOIV      ------AGCA CCTGTGTTAG TGTGGCCAAA A[illegible]ATATGA TCCGTGGTGA
 MODV       ------ACCA ATTGTGTGAC CATATCACAG AGAGAA[illegible] TTAAAGGAGT
 KRV        ---------T CTTCCATTGA GCCGTTGTAT ACCATCTCGG [illegible]TGGAAC

            ATCGGGAGCT ACGTGGGTTG ACGTGGTGCT CGAGCACGGT GGGTGTGTGA
            GTCAGGAGCT ACGTGGGTGG ATGTGGTACT GGAGCATGGA AGTTGCGTCA
            GTCAGGAGGG AGTTGGGTTG ACATAGTTTT AGAACATGGA AGTTGTGTGA
            CTCAGGTGGA GCATGGGTCG ACCTGGTGCT AGAACATGGA GGATGCGTCA
            GTCCGGAGCA ACATGGGTGG ACTTGGTCCT TGAAGGTGAC AGCTGTGTGA
            GTCTGGAGCT ACATGGGTTG ATCTGGTACT GGAAGGCGAT AGTTGTGTGA
            CAGTGGAGCC ACTTGGGTGG ACTTGGTGCT AGAAGGAGAT AGCTGCTTGA
            TTCAGGAGCT ACATGGGTCG ATTTGGTGCT GGAGGGCGAC AGTTGCATCA
```

SEQBOOT

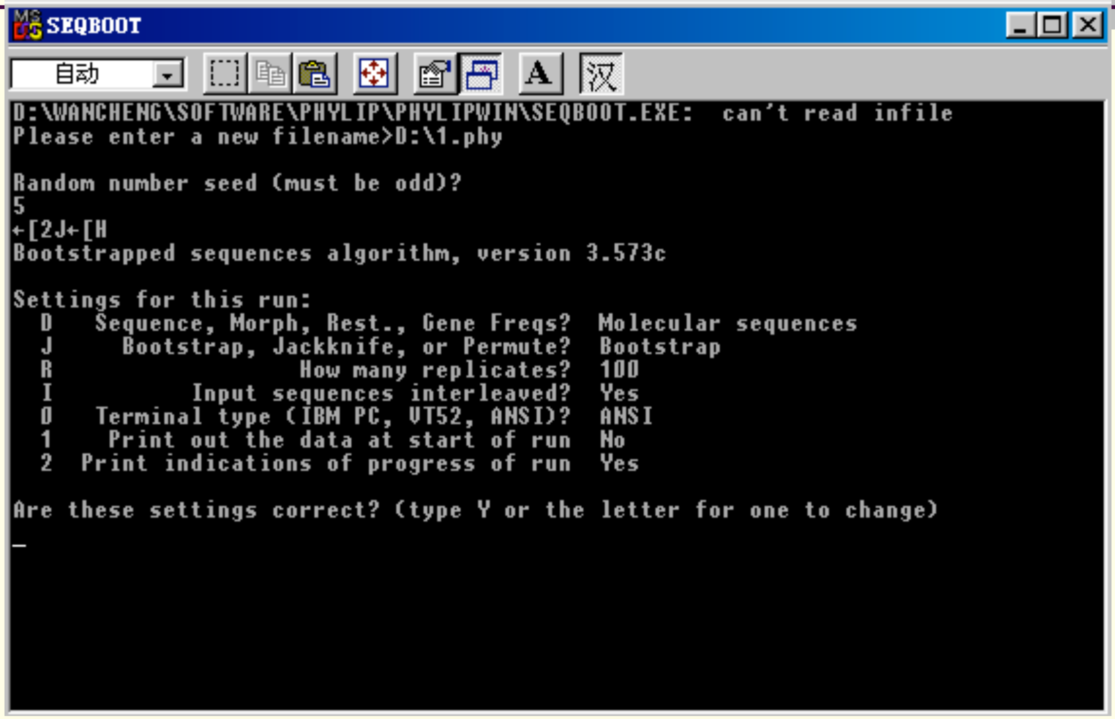


```
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\SEQBOOT.EXE:  can't read infile
Please enter a new filename>
```

1. The name of *.PHY

2. Input a Random number seed (must be odd)

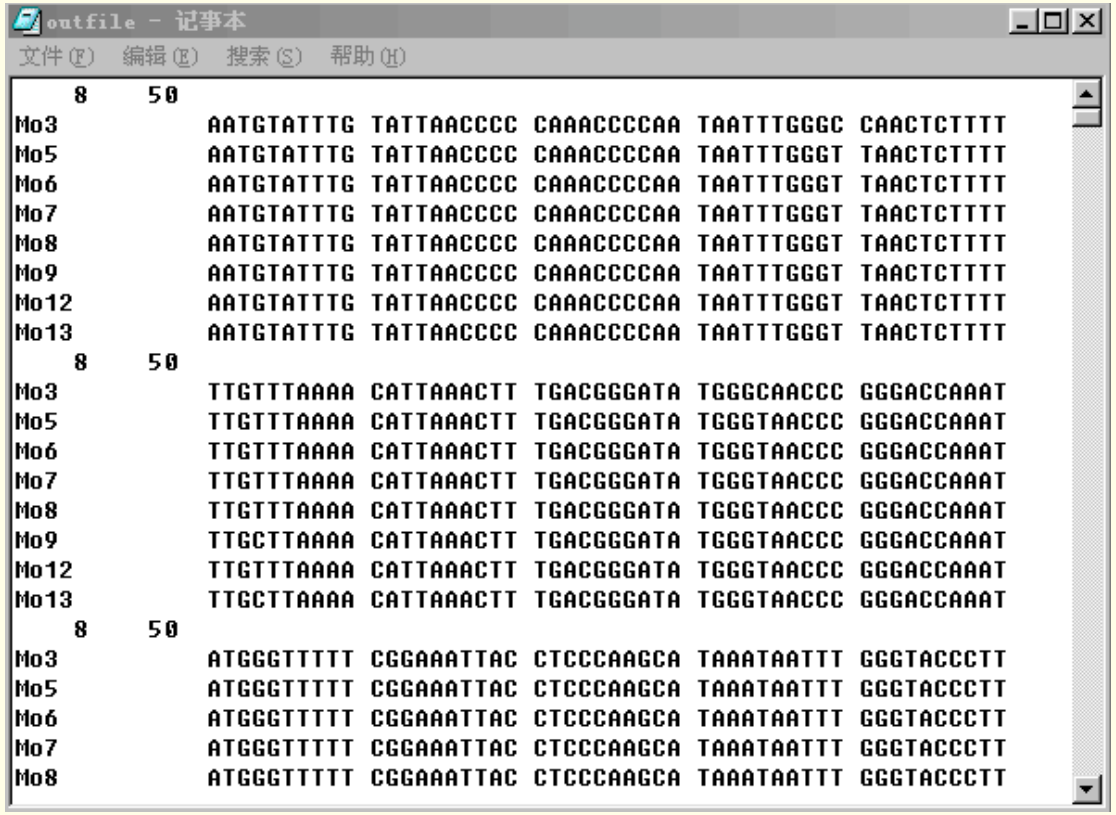
SEQBOOT



```
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\SEQBOOT.EXE:  can't read infile
Please enter a new filename>D:\1.phy

Random number seed (must be odd)?
5
←[2J←[H
Bootstrapped sequences algorithm, version 3.573c

Settings for this run:
  D    Sequence, Morph, Rest., Gene Freqs?  Molecular sequences
  J      Bootstrap, Jackknife, or Permute?  Bootstrap
  R                How many replicates?  100
  I          Input sequences interleaved?  Yes
  0    Terminal type (IBM PC, VT52, ANSI)?  ANSI
  1     Print out the data at start of run  No
  2   Print indications of progress of run  Yes

Are these settings correct? (type Y or the letter for one to change)
_
```

J == Bootstrap

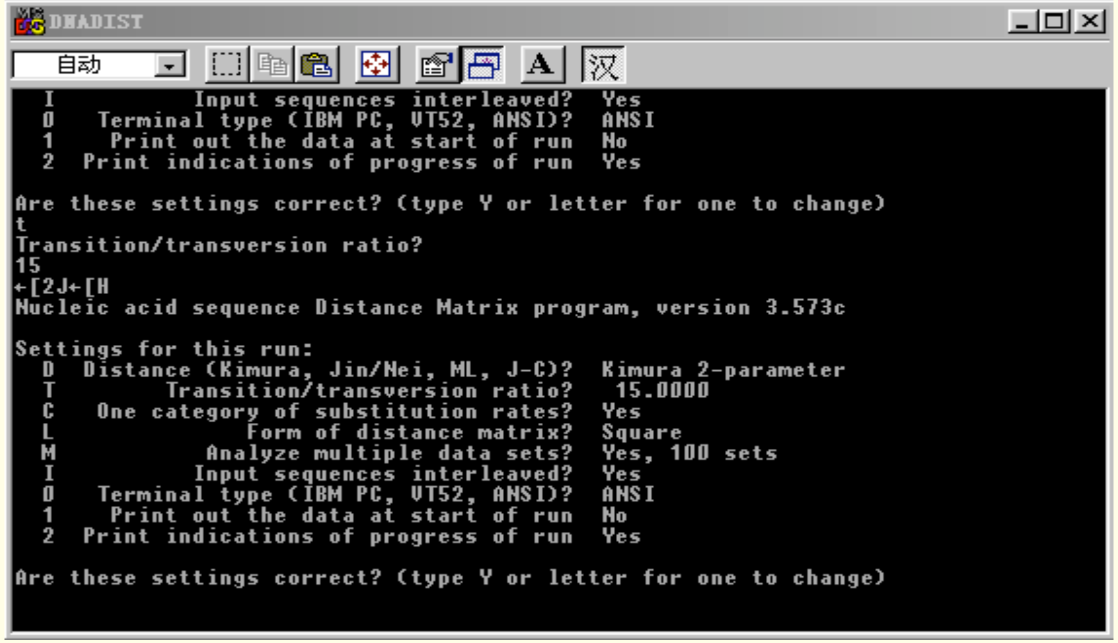
R == number of republicate, typical 100



```
    8    50
Mo3           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGC CAACTCTTTT
Mo5           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo6           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo7           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo8           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo9           AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo12          AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
Mo13          AATGTATTTG TATTAACCCC CAAACCCCAA TAATTTGGGT TAACTCTTTT
    8    50
Mo3           TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGCAACCC GGGACCAAAT
Mo5           TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo6           TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo7           TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo8           TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo9           TTGCTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo12          TTGTTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
Mo13          TTGCTTAAAA CATTAAACTT TGACGGGATA TGGGTAACCC GGGACCAAAT
    8    50
Mo3           ATGGGTTTTT CGGAAATTAC CTCCCAAGCA TAAATAATTT GGGTACCCTT
Mo5           ATGGGTTTTT CGGAAATTAC CTCCCAAGCA TAAATAATTT GGGTACCCTT
Mo6           ATGGGTTTTT CGGAAATTAC CTCCCAAGCA TAAATAATTT GGGTACCCTT
Mo7           ATGGGTTTTT CGGAAATTAC CTCCCAAGCA TAAATAATTT GGGTACCCTT
Mo8           ATGGGTTTTT CGGAAATTAC CTCCCAAGCA TAAATAATTT GGGTACCCTT
```

The result file with 100 replicate

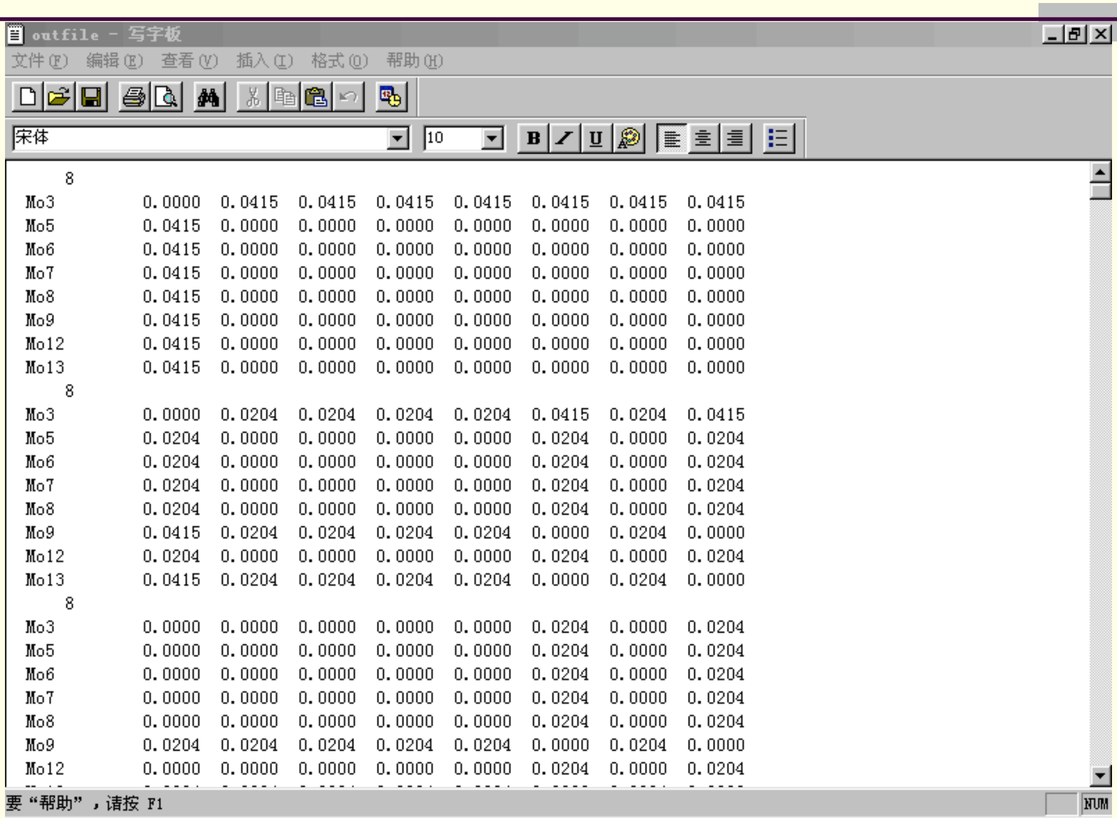
DNADIST



```
   I          Input sequences interleaved?  Yes
   0    Terminal type (IBM PC, VT52, ANSI)?  ANSI
   1     Print out the data at start of run  No
   2   Print indications of progress of run  Yes

Are these settings correct? (type Y or letter for one to change)
t
Transition/transversion ratio?
15
←[2J←[H
Nucleic acid sequence Distance Matrix program, version 3.573c

Settings for this run:
   D   Distance (Kimura, Jin/Nei, ML, J-C)?  Kimura 2-parameter
   T         Transition/transversion ratio?   15.0000
   C    One category of substitution rates?  Yes
   L               Form of distance matrix?  Square
   M            Analyze multiple data sets?  Yes, 100 sets
   I          Input sequences interleaved?  Yes
   0    Terminal type (IBM PC, VT52, ANSI)?  ANSI
   1     Print out the data at start of run  No
   2   Print indications of progress of run  Yes

Are these settings correct? (type Y or letter for one to change)
```

T: 15 ~ 30
M: 100

Distance Matrix

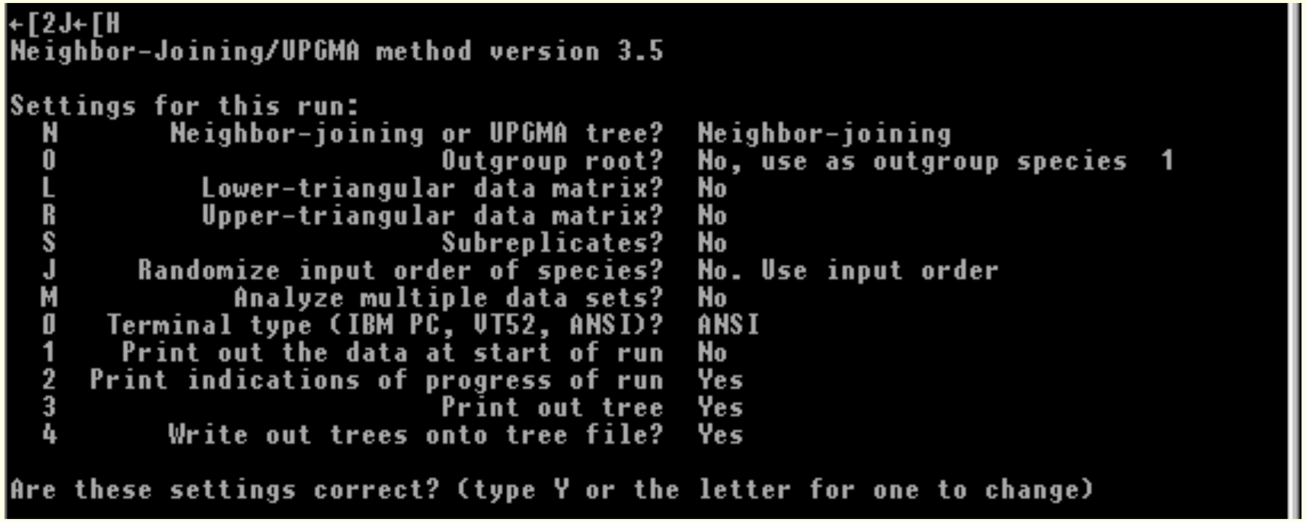


```
      8
Mo3         0.0000  0.0415  0.0415  0.0415  0.0415  0.0415  0.0415  0.0415
Mo5         0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo6         0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo7         0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo8         0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo9         0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo12        0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
Mo13        0.0415  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000  0.0000
      8
Mo3         0.0000  0.0204  0.0204  0.0204  0.0204  0.0415  0.0204  0.0415
Mo5         0.0204  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo6         0.0204  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo7         0.0204  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo8         0.0204  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo9         0.0415  0.0204  0.0204  0.0204  0.0204  0.0000  0.0204  0.0000
Mo12        0.0204  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo13        0.0415  0.0204  0.0204  0.0204  0.0204  0.0000  0.0204  0.0000
      8
Mo3         0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo5         0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo6         0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo7         0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo8         0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
Mo9         0.0204  0.0204  0.0204  0.0204  0.0204  0.0000  0.0204  0.0000
Mo12        0.0000  0.0000  0.0000  0.0000  0.0000  0.0204  0.0000  0.0204
```

要“帮助”，请按 F1

100 replica ➔ 100 distance matrix

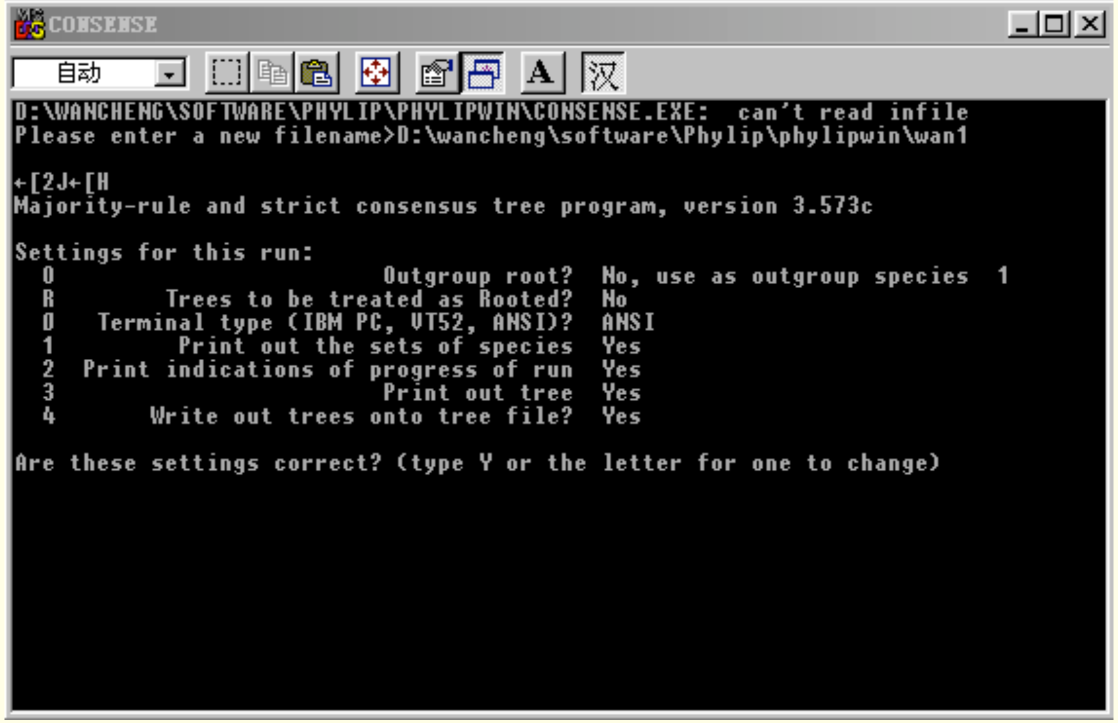
NEIGHBOR



```
←[2J←[H
Neighbor-Joining/UPGMA method version 3.5

Settings for this run:
  N       Neighbor-joining or UPGMA tree?  Neighbor-joining
  O                        Outgroup root?  No, use as outgroup species  1
  L         Lower-triangular data matrix?  No
  R         Upper-triangular data matrix?  No
  S                        Subreplicates?  No
  J     Randomize input order of species?  No. Use input order
  M           Analyze multiple data sets?  No
  0   Terminal type (IBM PC, VT52, ANSI)?  ANSI
  1    Print out the data at start of run  No
  2  Print indications of progress of run  Yes
  3                        Print out tree  Yes
  4       Write out trees onto tree file?  Yes

Are these settings correct? (type Y or the letter for one to change)
```

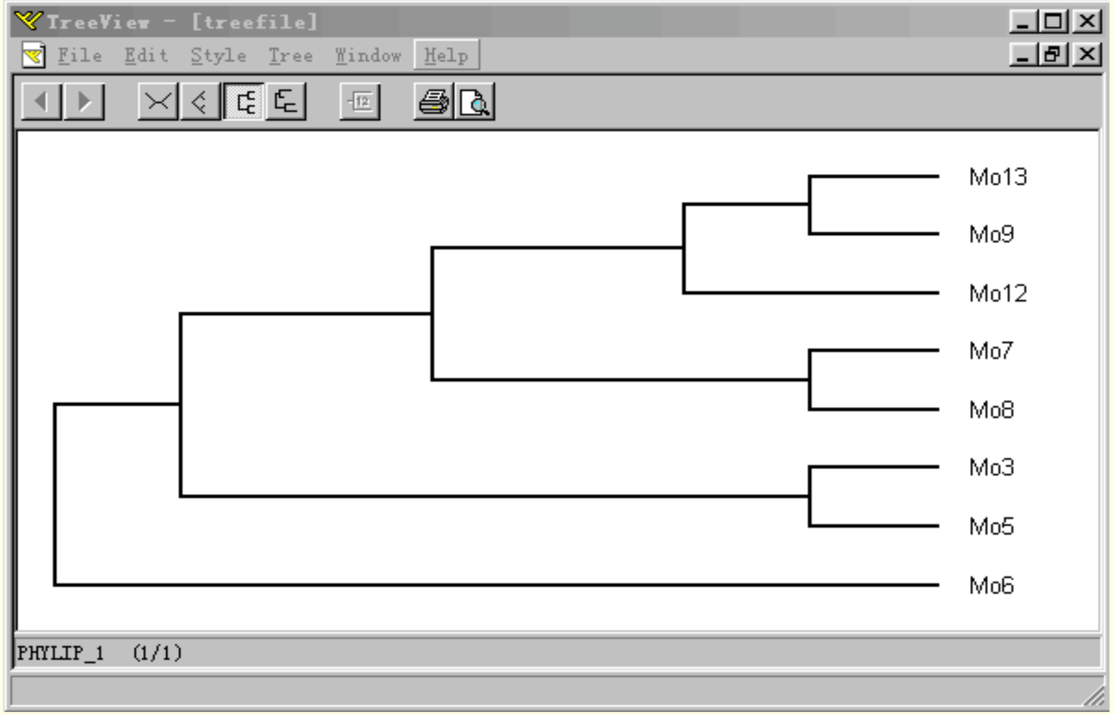
- M == 100

CONSENSE



```
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\CONSENSE.EXE:  can’t read infile
Please enter a new filename>D:\wancheng\software\Phylip\phylipwin\wan1

←[2J←[H
Majority-rule and strict consensus tree program, version 3.573c

Settings for this run:
  O                       Outgroup root?  No, use as outgroup species  1
  R         Trees to be treated as Rooted?  No
  D   Terminal type (IBM PC, VT52, ANSI)?  ANSI
  1          Print out the sets of species  Yes
  2  Print indications of progress of run  Yes
  3                       Print out tree  Yes
  4       Write out trees onto tree file?  Yes

Are these settings correct? (type Y or the letter for one to change)
```

View the Treefile by TREEVIEW



More Help on PHYLIP

- Homepage:
  - http://evolution.genetics.washington.edu/phylip.html

- A pretty good tutorial:
  - http://koti.mbnet.fi/tuimala/oppaat/phylip2.pdf

Thank you for your attentions!