Phylogeny

Phylogeny is the description of biological relationships, usually expressed as a tree.

A statement of phylogeny among objects assumes homology and depends on classification.

Phylogenetic analysis is an investigation of the evolutionary relationships among a group of related sequences by producing a tree representation of the relationships.

In fact, phylogenetic relationships among many kinds of organisms are difficult to determine in any other way.

Simply, organisms with high degrees of molecular similarity are expected to be closely related than those that are dissimilar.

Due to the availability of molecular data, taxonomists are forced to rely on comparisons of phenotypes (how organisms looked) to infer their genotypes (the genes that gave rise to their physical appearance).

Humans, flies, mollusks: light detecting organ-eye

Protein/DNA sequences

Phylogenetic trees

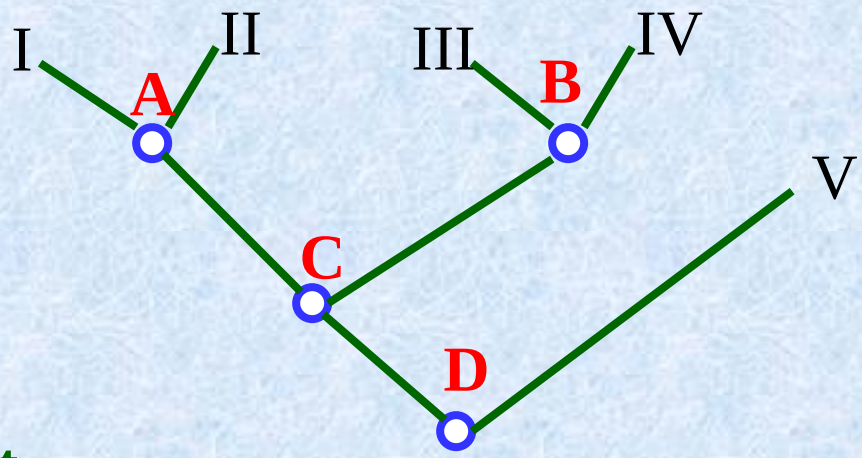
Graphical representation of the evolutionary relationship among three or more genes or organisms.

Relatedness of data; divergence times; nature of common ancestors

Nodes and branches:

Terminal nodes: at the tip of the branches (gene or organism: available data)

Internal node: common ancestor (data are not available)



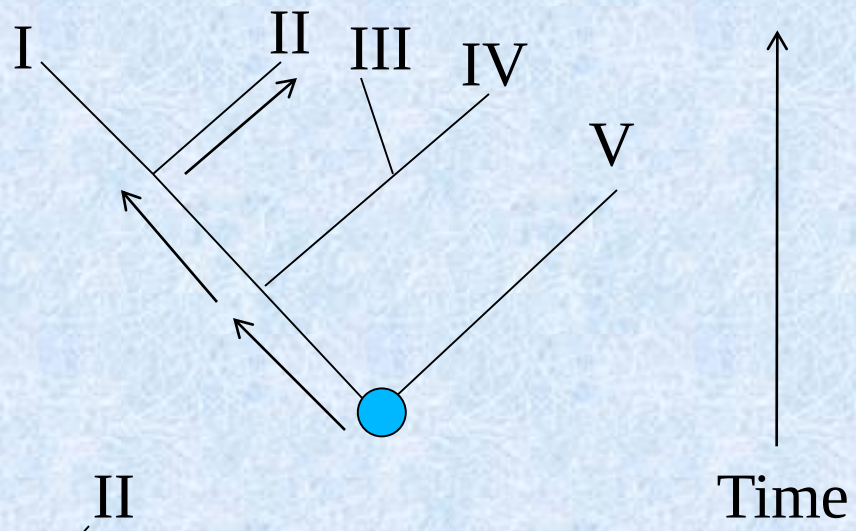
Bifurcating
Multifurcating

Scaled trees
Unscaled trees

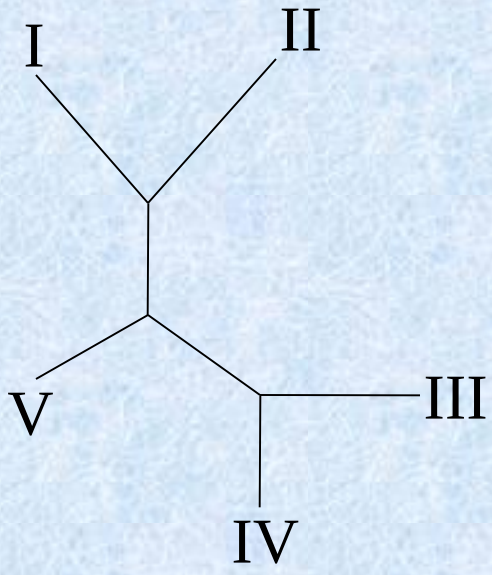
Newick format
(((I,II),(III,IV)), V)

Rooted and Unrooted trees

Rooted trees: a single node is designated as a common ancestor and a unique path leads it.

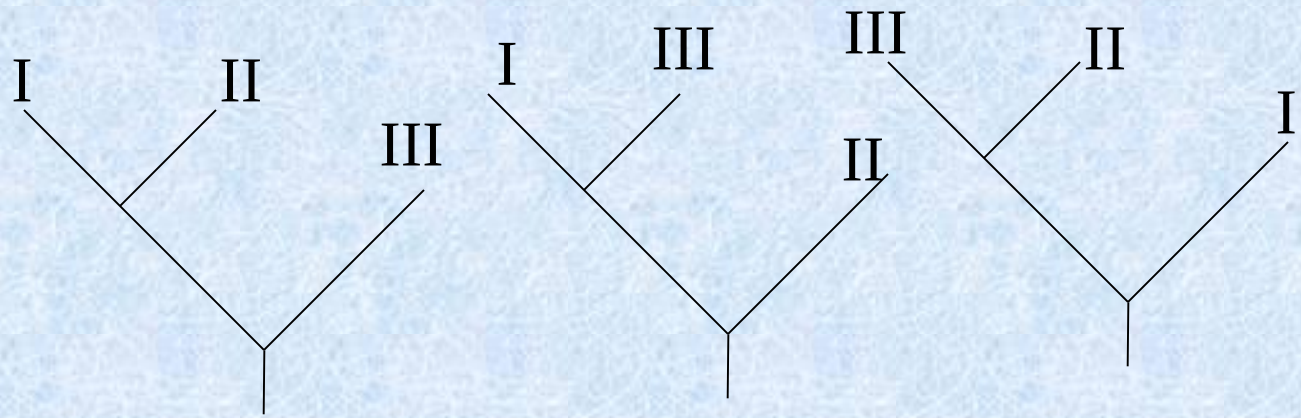


Unrooted trees: only relationship and not direction

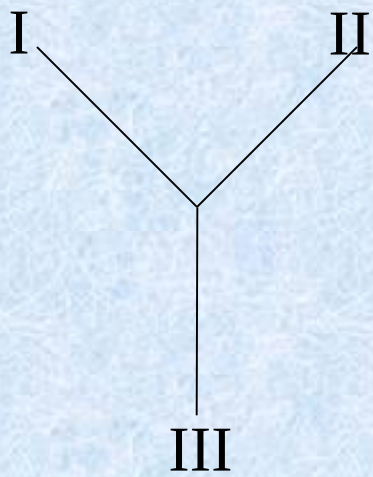


Possible rooted and unrooted trees
3 species

Rooted: 3



Unrooted: 1



$N_R = (2n-3)!/2^{n-2}(n-2)!$

$N_U=(2n-5)!/2^{n-3}(n-3)!$

2 data: Rooted: 1; Unrooted:1

3 data: R: 3, U: 1

4 data: R: 15, U: 3

Tree construction

1. UPGMA (Unweighted Pair Group Method with Arithmetic mean
2. Transformed Distance Method
3. Neighbor's Relation Method
4. Neighbor Joining Methods
5. Maximum Likelihood Approaches

UPGMA

Statistically based method

Requires data that can be condensed to genetic distance (distance matrix)

E.g. Species, A, B, C and D

Species	A	B	C
B	d_{AB}	-	-
C	d_{AC}	d_{BC}	-
D	d_{AD}	d_{BD}	d_{CD}

d_{AB}: the number of mismatching nucleotides (divided by total number of sites, where matches could have been found)

If A and B are related (minimum mismatches)

Form a group (AB)

Combine with other species, C and D

$$d_{(AB),C)} = 1/2\ (d_{AC} + d_{BC})$$

$$d_{(AB),D} = 1/2\ (d_{AD} + d_{BD})$$

The species separated by the smallest distance in the new matrix can be clustered together.

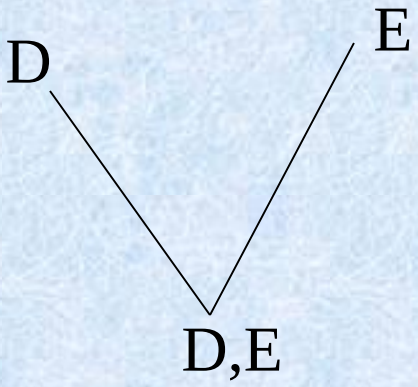
For scaled branches, the distances will be averaged.

Example

```
A: GTGCTGCACG GCTCAGTATA GCATTTACCC TTCCATCTTC AGATCCTGAA
B: ACGCTGCACG GCTCAGTGCG GTGCTTACCC TCCCATCTTC AGATCCTGAA
C: GTGCTGCACG GCTCGGCGCA GCATTTACCC TCCCATCTTC AGATCCTATC
D: GTATCACACG ACTCAGCGCA GCATTTGCCC TCCCGTCTTC AGATCCTAAA
E: GTATCACATA GCTCAGCGCA GCATTTGCCC TCCCGTCTTC AGATCTAAAA
```

Pairwise comparison (distance matrix)

Species	A	B	C	D
B	9	-	-	-
C	8	11	-	-
D	12	15	10	-
E	15	18	13	5



Pairwise comparison (distance matrix)

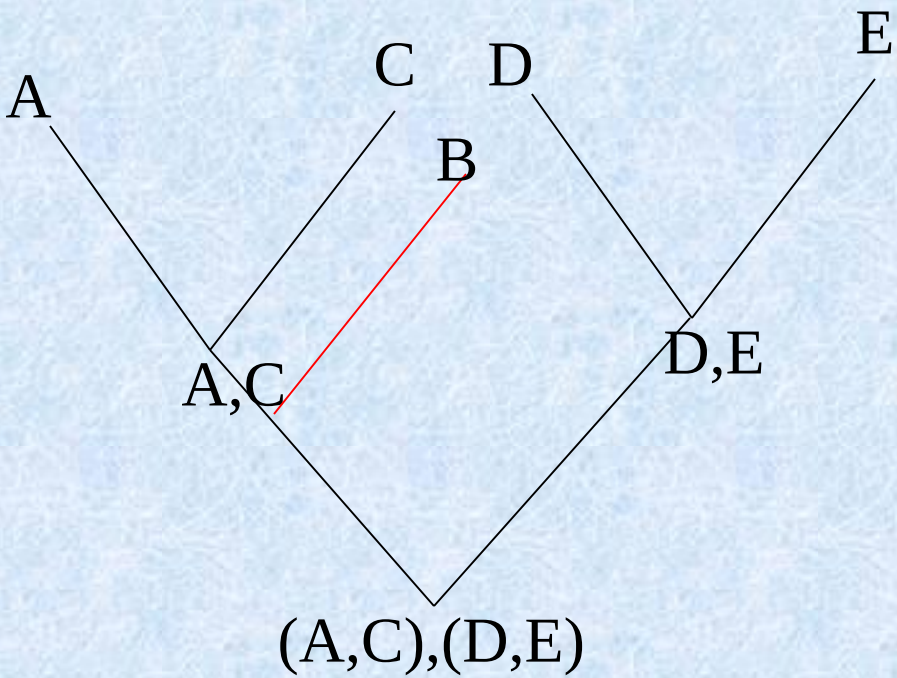
Species	A	B	C
B	9	-	-
C	8	11	-
DE	13.5	16.5	11.5

(AC),B = ½(AB+BC)

Pairwise comparison

Species	B	AC
AC	10	-
DE	16.5	12.5

(((A,C),B),D,E)



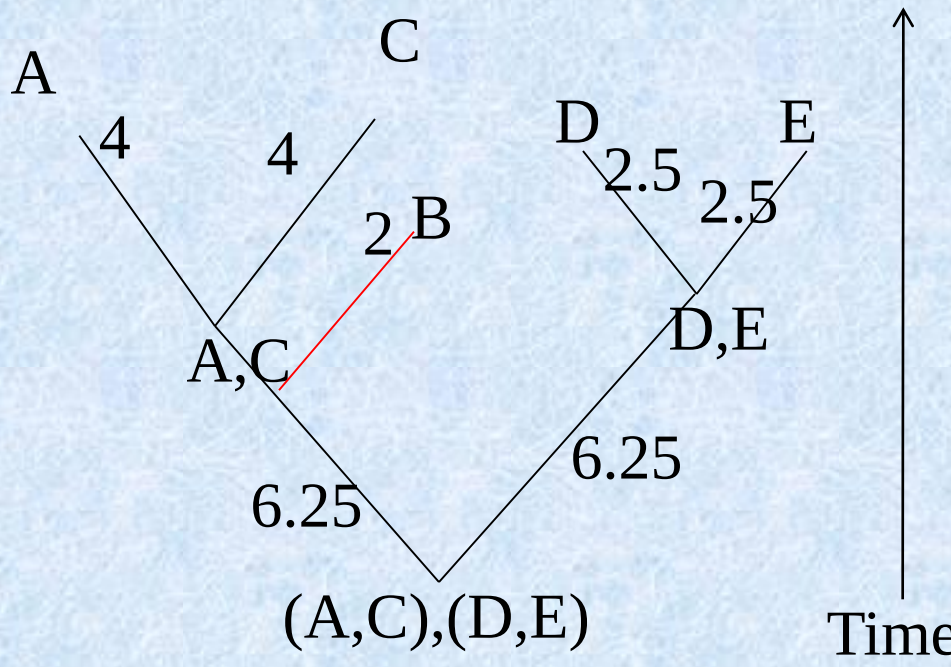
(((A,C),B),(D,E))

Estimation of branch lengths

Length of the branches can also be calculated with distance matrix

Pairwise comparison (distance matrix)

Species	A	B	C	D
B	9	-	-	-
C	8	11	-	-
D	12	15	10	-
E	15	18	13	5



Neighbor's Relation Method

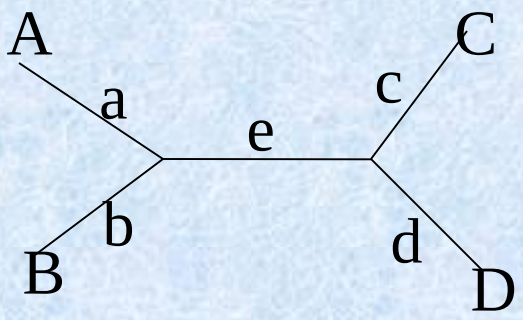
Another popular variant of UPGMA: tree is constructed with the smallest possible branch lengths overall.

Any unrooted tree, pairs of species that are separated from each other by just one internal node are said to be neighbors.

[UPGMA: Branch length is not additive.

E.g: d_{AE} = 4 + 6.25 + 6.25 + 2.5 = 19

Actual case: 15]



If additivity holds good:

$$d_{AC}+d_{BD} = d_{AD}+d_{BC} = a+b+c+d+2e = d_{AB}+d_{CD}+2e$$

a,b,c,d: lengths of terminal branches

e: length of central branch

Four point condition:

$d_{AB}+d_{CD} < d_{AC}+d_{BD}$

and

$d_{AB}+d_{CD} < d_{AD}+d_{BC}$

Considers all possible pairwise arrangements of four species and determines the arrangement, which satisfies the four point condition.

For four species, considers all possible values,

(i) $d_{AB}+d_{CD}$**; (ii)** $d_{AC}+d_{BD}$ **and (iii)** $d_{AD}+d_{BC}$

Smallest sum with pairing is 1 and others are 0

Repeat for all possible four pairs

Ones with highest scores are grouped.

New distance matrix can be generates as was done for UPGMA

Neighbor joining method

Tree is of star-like and all species comes as a single central node regardless of the number.

Difference with other methods is the way it determines the sum of branch lengths with each reiteration of the process.

$$S_{12} = (1/(2(N-2)) \Sigma(d_{1k}+d_{2k}) + (1/2)d_{12} + (1/N-2)\Sigma d_{ij}$$

Any pair of species can take positions 1 and 2; k is an accepted outgroup

Simplified into $Q_{12} = (N-2)d_{12} - \Sigma d_{1i} - \Sigma d_{2i}$

All possible pairs are considered and the pairs with smallest distance is taken.

Construct new distance matrix as done with UPGMA and repeat the process.

Maximum likelihood approaches

It represent an alternative and purely statistically based method of phylogenetic reconstruction.

Probabilities are considered for every individual nucleotide substitution.

Transitions (purine to purine/ pyramidine to pyramidine) and transversions

Multiple substitutions occurred at one or more sites, which are not necessarily independent.

It is necessary to take into account of all these facts, which needs heavy computational power.

With current facilities, it is possible to use the method for tree construction.

Program to construct trees

- **as Windows executables** (not counting executing in a "DOS box"). Programs available as source code which is Windows-specific are li
(Note that compilers available on Windows systems, particularly the free Cygwin and MinGW compilers, can also be used to compile ma
generic source code). Programs run in interpreted environments such as Perl, Python, R or MATLAB can also be run under Windows if t
programs are listed above under Unix.

PHYLIP	DNASIS	Mesquite	MrModeltest	MESA
PAUP*	MINSPNET	Phyledit	SymmeTREE	MultiPhyl
TREECON	BioEdit	SYN-TAX	TreeJuxtaposer	NimbleTree
GDA	ProSeq	PTP	Network	ArboDraw
SeqPup	PAL	DIVA	Spectronet	SPAGeDi
MOLPHY	WINCLADA	TreeFitter	Phylogen	CBCAnalyzer
GeneDoc	NONA	Phylo_win	Phylap	DualBrothers
COMPONENT	Phylogenetic Independence	SplitsTree	Dnatree	PaupUp
TREEMAP	PEBBLE	PAST	IMa2	Notung
COMPARE	HY-PHY	GeneStudio Pro	ProtTest	SSA
RAPDistance	TreeExplorer	Treefinder	GEODIS	Multidivtime
TreeView	Genie	PPH	TreeSetViz	ParaFit
Phylodendron	Vanilla	MetaPIGA	TreeMe	IDC
POPGENE	MEGA	Phyltools	ModelGenerator	TreeMaker
TFPGA	TNT	MSA	Simplot	CodonRates
GeneTree	GelCompar II	Mgenome	PHYLOGR	BAli-Phy
MVSP	Bionumerics	APE	ProfDist	CoMET
RSTCALC	TCS	PHASE	START2	TreeDyn
Genetix	FORESTER	PHYML	IQPNNI	DigTree
NJplot	Populations	YCDMA	STC	Geneious
unrooted	T-REX	NSA	TreeSAAP	Brownie
Arlequin	MrBayes	BEAST	Swaap	Mac5
DAMBE	EDIBLE	Clann	Swaap PH	BayesPhylogenies
DnaSP	Winboot	Jevtrace	TreeGraph 2	BayesTraits
PAML	r8s	MrMTgui	DIVERGE	MrEnt

Phylip

Phylip

Phylip is a program to create phylogenetic tree for a given set of amino acid sequences.

It takes the multiple alignment of the sequences as input

Multiple sequence alignments can be done with ClustalW, MAFFT etc.

MAFFT is widely used to prepare the input multiple alignment file suitable for Phylip

MAFFT

mafft.cbrc.jp/alignment/software/

MAFFT version 6

Multiple alignment program for amino acid or nucleotide sequences

CBRC AIST

Download version
- Mac OS X
- Windows
- Linux
- Source

Online version
- Alignment
- Phylogeny
- Rough tree

Merits / limitations
Algorithms
Tips
Benchmarks
Feedback

About

MAFFT is a multiple sequence alignment program for unix-like operating systems. It offers a range of multiple alignment methods, L-INS-i (accurate; for alignment of <~200 sequences), FFT-NS-2 (fast; for alignment of <~10,000 sequences), *etc.*

Download and Installation

- **Mac OS X**
- **Linux**
- **Windows**
- **Source**
- Changelog

The latest version is 6.857. ***New!*** (2011/05/30)

Input Format

Fasta format. example1 (LSU rRNA), example2 (protein)

The type of input sequences (amino acid or nucleotide) is automatically recognized.

Usage

```
% mafft [arguments] input > output
```

MAFFT

UPPERCASE / lowercase:

○ Same as input

Parameters:

Scoring matrix for amino acid sequences: BLOSUM62

Scoring matrix for nucleotide sequences: 200PAM / κ=2

† Switch it to '1PAM / κ=2' when aligning closely related DNA sequences.

Gap opening penalty: 1.53 (1.0 – 3.0)

Offset value: 0.0 (0.0 – 1.0)

† If long gaps are not expected, set it as 0.1 or larger value.

Mafft-homologs (Collects homologs from SwissProt by BLAST and performs profile-based alignments; Protein only): Help

☐ On

☐ Show homologs (if any)

Number of homologs: 50 (5 – 200)

Threshold: *E*= 1e-10 (1e-5 – 1e-40)

Plot LAST hits (DNA only):

◉ The top sequence vs the others ○ The longest sequence vs the others

◉ Plot and alignment ○ Plot only ○ Alignment only

Threshold: score=39 (E=8.4e-11)

Submit Reset

with <200 sequences × <1,000 nucleotides) Help

MAFFT Results

Jalview

Reformat to GCG, PHYLIP, MSF, NEXUS, uppercase/lowercase, *etc.* with Readseq

Phylogenetic Tree

MAFFT-G-INS-i Result

```
CLUSTAL format alignment by MAFFT (

sp|P69905|HBA_H MVLSPADKTNVKAAWGKVG
sp|P69907|HBA_P MVLSPADKTNVKAAWGKVG
sp|P06635|HBA_P MVLSPADKTNVKTAWGKVG
sp|P01966|HBA_B MVLSAADKGNVKAAWGKVG
sp|P01958|HBA_H MVLSAADKTNVKAAWSKVG
sp|P01959|HBA_E MVLSAADKTNVKAAWSKVG
sp|P01942|HBA_M MVLSGEDKSNIKAAWGKIG
sp|P01946|HBA_R MVLSADDKTNIKNCWGKIG
sp|P01965|HBA_P -VLSAADKANVKAAWGKVG
sp|P60529|HBA_C -VLSPADKTNIKSTWDKIG
                 ***  ** *:*  *.*:*

sp|P69905|HBA_H KKVADALTNAVAHVDDMPN
sp|P69907|HBA_P KKVADALTNAVAHVDDMPN
sp|P06635|HBA_P KKVADALTNAVAHVDDMPN
sp|P01966|HBA_B AKVAAALTKAVEHLDDLPG
sp|P01958|HBA_H KKVGDALTLAVGHLDDLPG
sp|P01959|HBA_E KKVGDALTLAVGHLDDLPG
sp|P01942|HBA_M KKVADALASAAGHLDDLPG
sp|P01946|HBA_R KKVADALAKAADHVEDLPG
sp|P01965|HBA_P QKVADALTKAVGHLDDLPG
sp|P60529|HBA_C KKVADALTTAVAHLDDLPG
                 **. **: *. *::*:*.
```

Submit Reset

Options

Output sequence format:

Pretty

IG|Stanford
GenBank|gb
NBRF
EMBL|em
GCG
DNAStrider
Pearson|Fasta|fa
Phylip3.2
Phylip|Phylip4
Plain|Raw
PIR|CODATA
MSF
PAUP|NEXUS
Pretty
XML
Clustal
FlatFeat|FFF
GFF
ACEDB

☐ Remove gap symbols: -

☐ Calculate checksum of sequences

Select ◉ all, or ○ sequences by number:

☐ Translate bases (list as from-base:to-base pairs)

seq by D.G. Gilbert, 2.1.26 (18-Oct-2007)

tp://iubio.bio.indiana.edu/soft/molbio/readseq/java/

```
 10 142
sp|P69905|    MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS
sp|P69907|    MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS
sp|P06635|    MVLSPADKTN VKTAWGKVGA HAGDYGAEAL ERMFLSFPTT KTYFPHFDLS
sp|P01966|    MVLSAADKGN VKAAWGKVGG HAAEYGAEAL ERMFLSFPTT KTYFPHFDLS
sp|P01958|    MVLSAADKTN VKAAWSKVGG HAGEYGAEAL ERMFLGFPTT KTYFPHFDLS
sp|P01959|    MVLSAADKTN VKAAWSKVGG NAGEFGAEAL ERMF
sp|P01942|    MVLSGEDKSN IKAAWGKIGG HGAEYGAEAL ERMF
sp|P01946|    MVLSADDKTN IKNCWGKIGG HGGEYGEEAL QRMF
sp|P01965|    -VLSAADKAN VKAAWGKVGG QAGAHGAEAL ERMF
sp|P60529|    -VLSPADKTN IKSTWDKIGG HAGDYGGEAL DRTF

              HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSAL
              HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSAL
              HGSAQVKDHG KKVADALTNA VAHVDDMPNA LSAL
              HGSAQVKGHG AKVAAALTKA VEHLDDLPGA LSEL
              HGSAQVKAHG KKVGDALTLA VGHLDDLPGA LSNL
              HGSAQVKAHG KKVGDALTLA VGHLDDLPGA LSNL
              HGSAQVKGHG KKVADALASA AGHLDDLPGA LSAL
              PGSAQVKAHG KKVADALAKA ADHVEDLPGA LSTL
              HGSDQVKAHG QKVADALTKA VGHLDDLPGA LSAL
              PGSAQVKAHG KKVADALTTA VAHLDDLPGA LSAL

              LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVS
              LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVS
              LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
              LLSHSLLVTL ASHLPSDFTP AVHASLDKFL ANVSTVLTSK YR
              LLSHCLLSTL AVHLPNDFTP AVHASLDKFL SSVSTVLTSK YR
              LLSHCLLSTL AVHLPNDFTP AVHASLDKFL STVSTVLTSK YR
              LLSHCLLVTL ASHHPADFTP AVHASLDKFL ASVSTVLTSK YR
              FLSHCLLVTL ACHHPGDFTP AMHASLDKFL ASVSTVLTSK YR
              LLSHCLLVTL AAHHPDDFNP SVHASLDKFL ANVSTVLTSK YR
              LLSHCLLVTL ACHHPTEFTP AVHASLDKFF AAVSTVLTSK YR
```

Submit Reset

Options

Output sequence format:
Phylip|Phylip4

Return biosequence data:
- Download to file
- View in browser

Change sequence case to
- No change
- lower
- UPPER

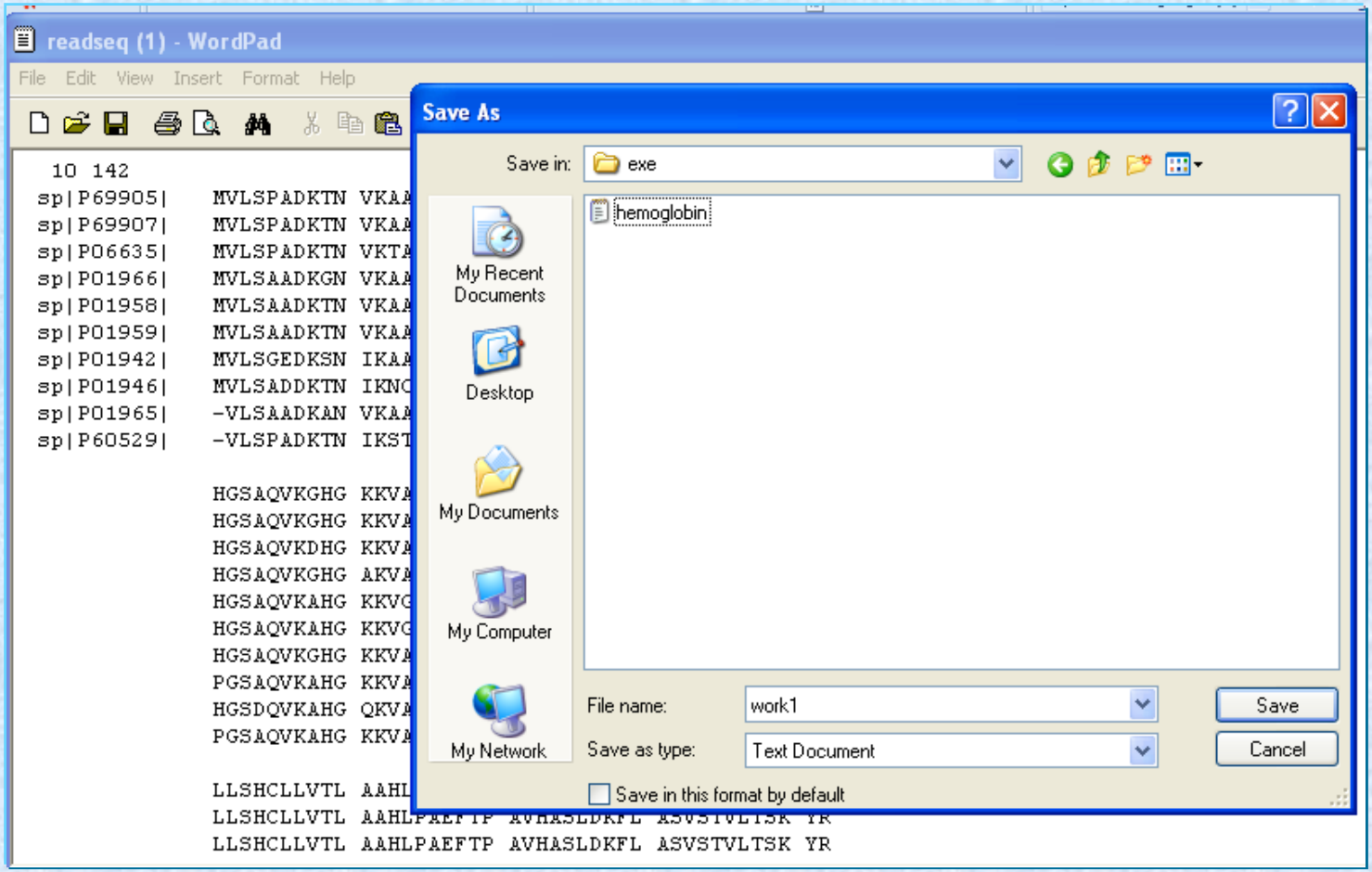
- Remove gap symbols: -
- Calculate checksum of sequences

Select all, or sequences by number:

- Translate bases (list as from-base:to-base pairs)

Saved in a temporary file "readseq".

Open it and save as "work1"



Folder: Phylip-3.69/exe/work1

Procedure to run Phylip

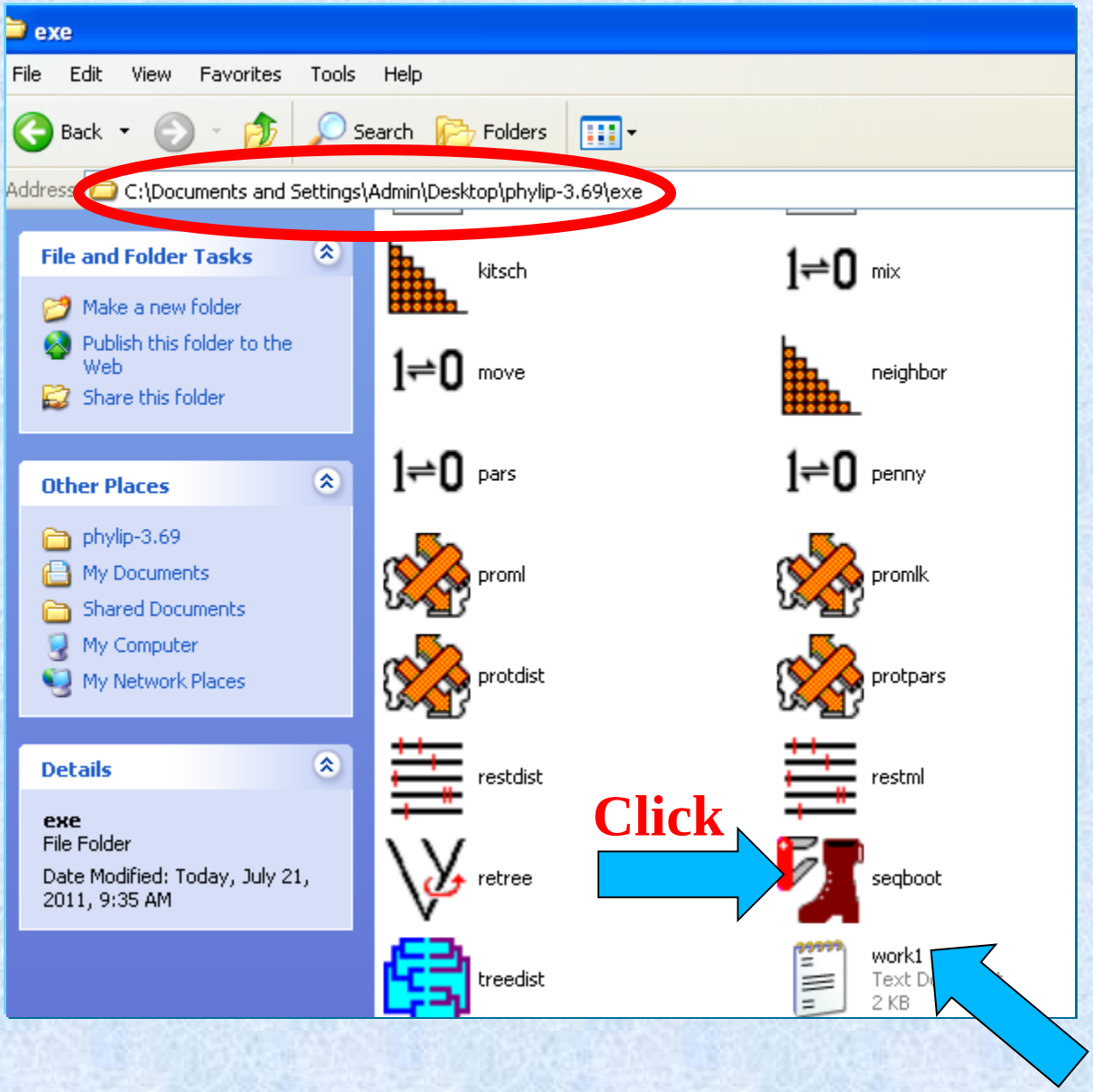
1. Bootstrapping: to check the confidence level

In statistics, bootstrapping is a computer-based method for assigning measures of accuracy to sample estimates.

Bootstrapping is the practice of estimating properties of an estimator (such as its variance) by measuring those properties when sampling from an approximating distribution.

One standard choice for an approximating distribution is the empirical distribution of the observed data.

This can be implemented by constructing a number of resamples of the observed dataset, each of which is obtained by random sampling with replacement from the original dataset.


exe
File Edit View Favorites Tools Help
Back
Search
Folders
Address
C:\Documents and Settings\Admin\Desktop\phylip-3.69\exe
File and Folder Tasks
Make a new folder
Publish this folder to the Web
Share this folder
Other Places
phylip-3.69
My Documents
Shared Documents
My Computer
My Network Places
Details
exe
File Folder
Date Modified: Today, July 21, 2011, 9:35 AM
kitsch
mix
move
neighbor
pars
penny
proml
promlk
protdist
protpars
restdist
restml
retree
seqboot
treedist
work1
2 KB
Click

```
C:\Documents and Settings\Admin\Desktop\phylip-3.69\exe\seqboot.exe
seqboot.exe: can't find input file "infile"
Please enter a new file name> work1.txt
```

```
Bootstrapping algorithm, version 3.69

Settings for this run:
  D       Sequence, Morph, Rest., Gene Freqs?  Molecular sequences
  J   Bootstrap, Jackknife, Permute, Rewrite?  Bootstrap
  %     Regular or altered sampling fraction?  regular
  B       Block size for block-bootstrapping?  1 (regular bootstrap)
  R                    How many replicates?    100
  W               Read weights of characters?  No
  C                 Read categories of sites?  No
  S      Write out data sets or just weights?  Data sets
  I              Input sequences interleaved?  Yes
  0       Terminal type (IBM PC, ANSI, none)?  IBM PC
  1        Print out the data at start of run  No
  2      Print indications of progress of run  Yes

  Y to accept these or type the letter for one to change
```

```
completed replicate number    1
completed replicate number    2
completed replicate number    3
completed replicate number    4
completed replicate number    5
completed replicate number    6
completed replicate number    7
completed replicate number    8
completed replicate number    9
completed replicate number   10

Output written to file "outfile"

Done.

Press enter to quit.
```

```
Bootstrapping algorithm, version 3.69

Settings for this run:
  D       Sequence, Morph, Rest., Gene Freqs?  Molecular s
  J   Bootstrap, Jackknife, Permute, Rewrite?  Bootstrap
  %     Regular or altered sampling fraction?  regular
  B       Block size for block-bootstrapping?  1 (regular bootstrap)
  R                    How many replicates?    10
  W               Read weights of characters?  No
  C                 Read categories of sites?  No
  S      Write out data sets or just weights?  Data sets
  I              Input sequences interleaved?  Yes
  0       Terminal type (IBM PC, ANSI, none)?  IBM PC
  1        Print out the data at start of run  No
  2      Print indications of progress of run  Yes

  Y to accept these or type the letter for one to change
R
Number of replicates?
10
```

outfile contains 10 replications

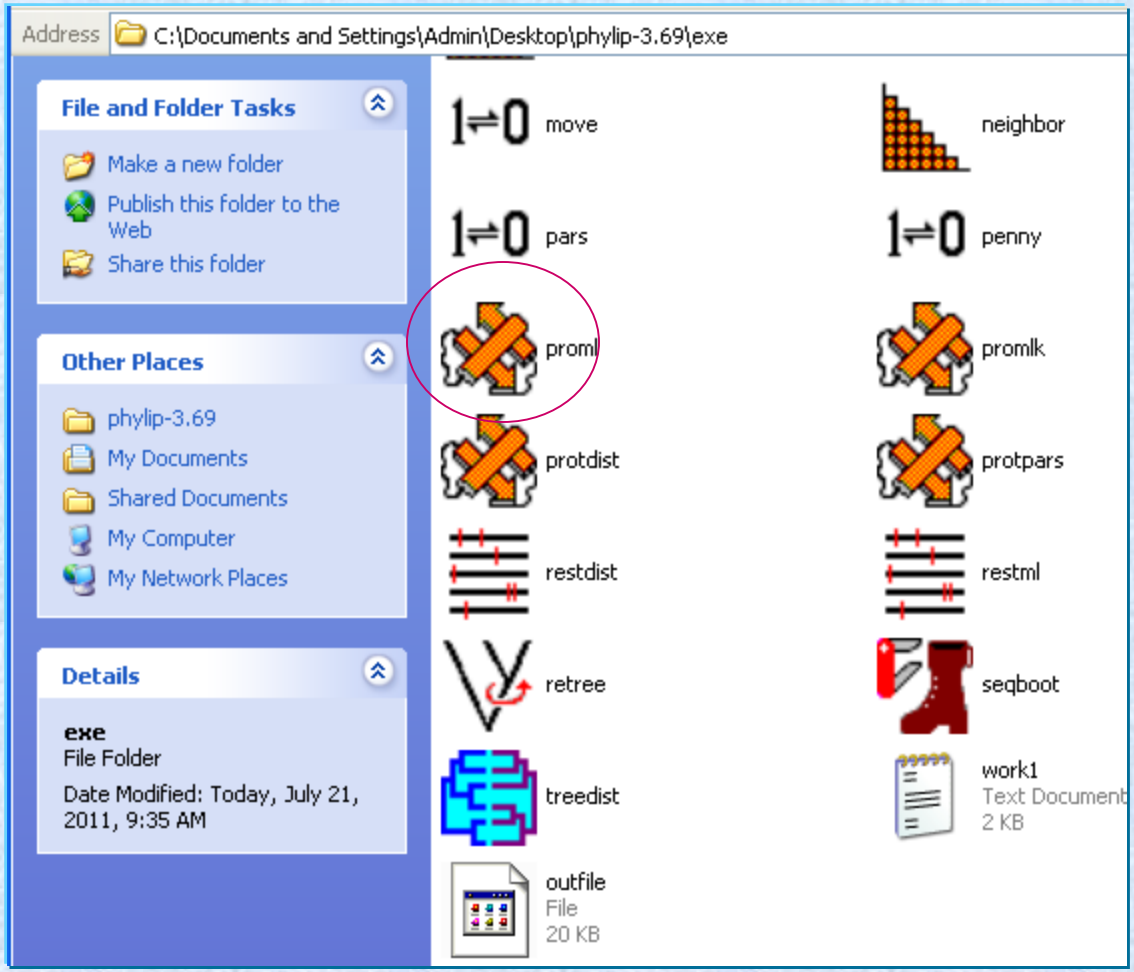
```
  Y to accept these or type the letter for one to change
y

Random number seed (must be odd)?
5
```

```
   10   142
sp|P69905| MLPPADKKTT TVVAGKGAHH HGGEGAELRM MMMFFSTTKK KTYFFFFDDS SHHGGAAKGK
sp|P69907| MLPPADKKTT TVVAGKGAHH HGGEGAELRM MMMFFSTTKK KTYFFFFDDS SHHGGAAKGK
sp|P06635| MLPPADKKTT TVVTGKGAHH HGGDGAELRM MMMFFSTTKK KTYFFFFDDS SHHGGAAKGK
sp|P01966| MLAAADKKGG GVVAGKGGHH HAAEGAELRM MMMFFSTTKK KTYFFFFDDS SHHGGAAKGA
sp|P01958| MLAAADKKTT TVVASKGGHH HGGEGAELRM MMMFFGTTKK KTYFFFFDDS SHHGGAAKGK
sp|P01959| MLAAADKKTT TVVASKGGNN NGGEGAELRM MMMFFGTTKK KTYFFFFDDS SHHGGAAKGK
sp|P01942| MLGGEDKKSS SIIAGKGGHH HAAEGAELRM MMMFFSTTKK KTYFFFFDDS SHHGGAAKGK
sp|P01946| MLAADDKKTT TIINGKGGHH HGGEGEELRM MMMFFATTKK KTYIIIIDDS SPPGGAAKGK
sp|P01965| -LAAADKKAA AVVAGKGGQQ QGGAGAELRM MMMFFGTTKK KTYFFFFNNS SHHGGDDKGQ
sp|P60529| -LPPADKKTT TIISDKGGHH HGGDGGELRT TTTFFSTTKK KTYFFFFDDS SPPGGAAKGK

           KKKVADDDLA AAADPSALLH HAHHRRDDNN KLLLHCCCCV TTTLLAAAHL LAAAEFFTHA
           KKKVADDDLA AAADPSALLH HAHHRRDDNN KLLLHCCCCV TTTLLAAAHL LAAAEFFTHA
           KKKVADDDLA AAADPSALLH HAHHRRDDNN KLLLHCCCCV TTTLLAAAHL LAAAEFFTHA
           AKKVAAAALA AAEDPSELLH HAHHRRDDNN KLLLHSSSSV TTTLLAASHL LSSSDFFTHA
           KKKVGDDDLA AAGDPSNLLH HAHHRRDDNN KLLLHCCCCS TTTLLAAVHL LNNNDFFTHA
           KKKVGDDDLA AAGDPSNLLH HAHHRRDDNN KLLLHCCCCS TTTLLAAVHL LNNNDFFTHA
           KKKVADDDLA AAGDPSALLH HAHHRRDDNN KLLLHCCCCV TTTLLAASHH HAAADFFTHA
           KKKVADDDLA AADEPSTLLH HAHHRRDDNN KFFFHCCCCV TTTLLAACHH HGGGDFFTHA
           QKKVADDDLA AAGDPSALLH HAHHRRDDNN KLLLHCCCCV TTTLLAAAHH HDDDDFFNHA
           KKKVADDDLA AAADPSALLH HAYYRRDDNN KLLLHCCCCV TTTLLAACHH HTTTEFFTHA

           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR
           ASSLLFFVVS STLLSKYYYR RR

   10   142
sp|P69905| MLAAAWVGGA AGEYYGGAAL EEERFFLLLF FYFFFFPPHD DDLLHAAQQQ QVVVKGGKKA
sp|P69907| MLAAAWVGGA AGEYYGGAAL EEERFFLLLF FYFFFFPPHD DDLLHAAQQQ QVVVKGGKKA
sp|P06635| MLAATWVGGA AGDYYGGAAL EEERFFLLLF FYFFFFPPHD DDLLHAAQQQ QVVVKDGKKA
sp|P01966| MLAAAWVGGA AAEYYGGAAL EEERFFLLLF FYFFFFPPHD DDLLHAAQQQ QVVVKGGKKA
sp|P01958| MLAAAWVGGA AGEYYGGAAL EEERFFLLLF FYFFFFPPHD DDLLHAAQQQ QVVVKAGKKG
```

outfile
10 different sets

Phylogenetic tree using Maximum likelihood method

The program is proml



outfile obtained from seqboot is the input for proml

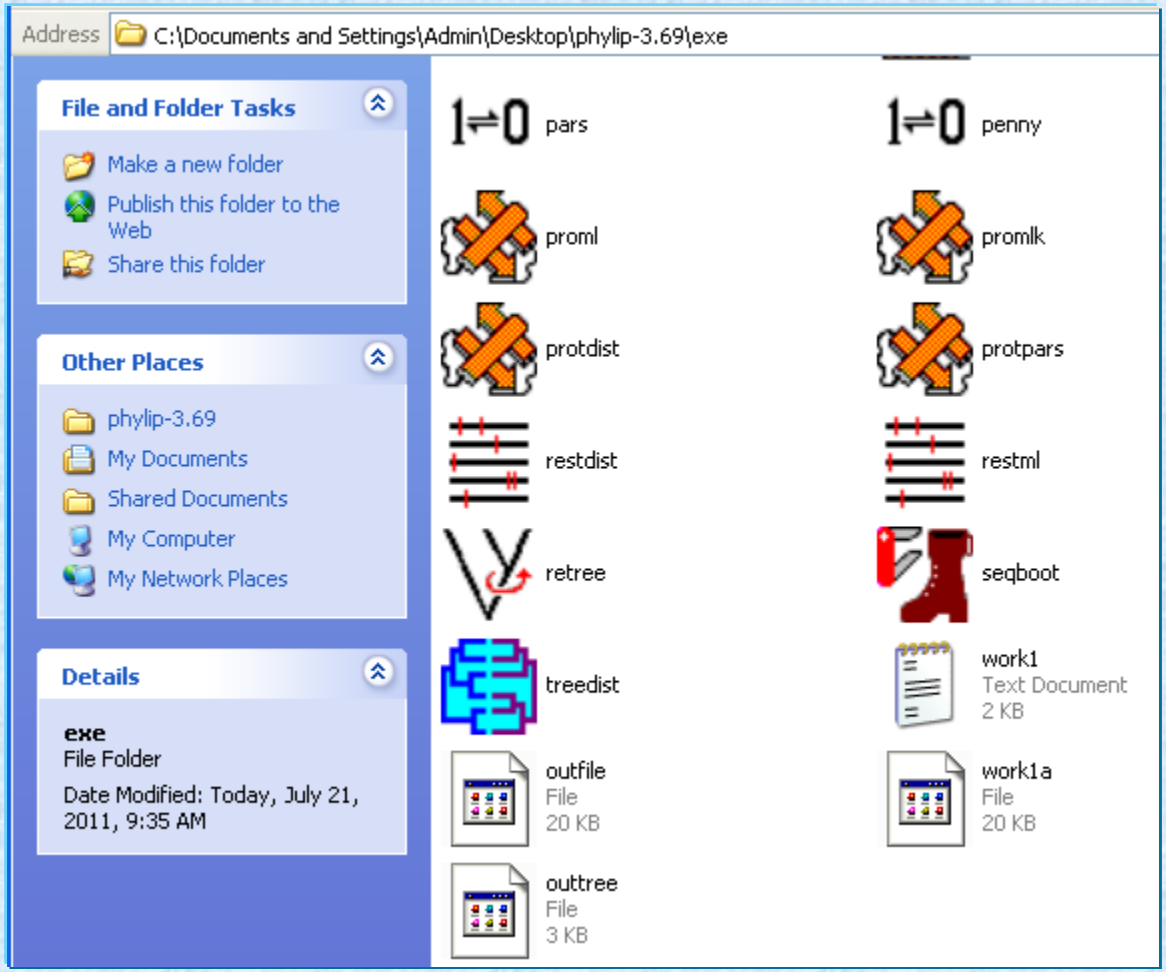
```
C:\Documents and Settings\Admin\Desktop\phylip-3.69\exe\proml.exe
proml.exe: can't find input file "infil
Please enter a new file name> outfile

proml.exe: the file "outfile" that you wanted to
     use as output file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
F
Please enter a new file name> work1a
```

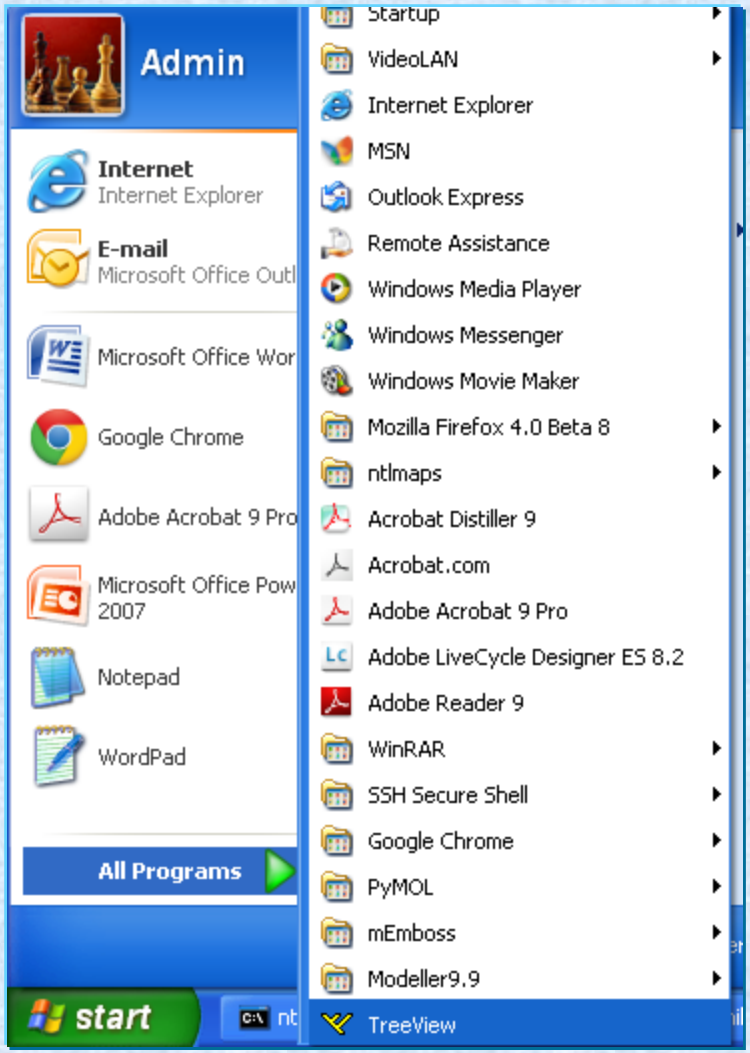
```
Amino acid sequence Maximum Likelihood method, version 3.69

Settings for this run:
  U                 Search for best tree?  Yes
  P    JTT, PMB or PAM probability model?  Jones-Taylor-Thornton
  C               One category of sites?  Yes
  R           Rate variation among sites?  constant rate of change
  W                      Sites weighted?  No
  S        Speedier but rougher analysis?  Yes
  G               Global rearrangements?  No
  J   Randomize input order of sequences?  No. Use input order
  O                       Outgroup root?  No, use as outgroup species  1
  M           Analyze multiple data sets?  No
  I          Input sequences interleaved?  Yes
  0   Terminal type (IBM PC, ANSI, none)?  IBM PC
  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
  5   Reconstruct hypothetical sequences?  No

  Y to accept these or type the letter for one to change
m
Multiple data sets or multiple weights? (type D or W)
d
How many data sets?
10
```

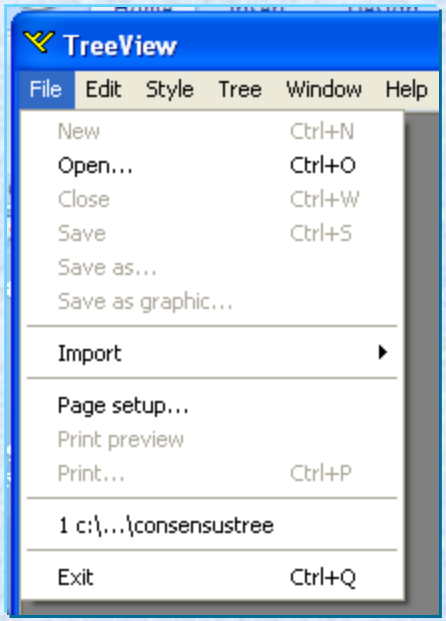
```
Random number seed (must be odd)?
5
Number of times to jumble?
3
```

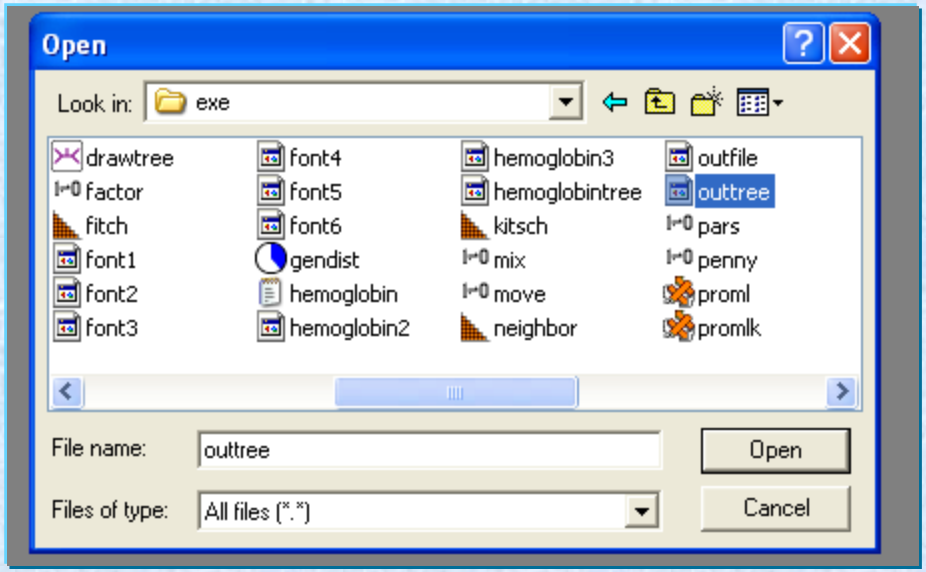
```
  Y to accept these or type the letter for one to change
Y
```

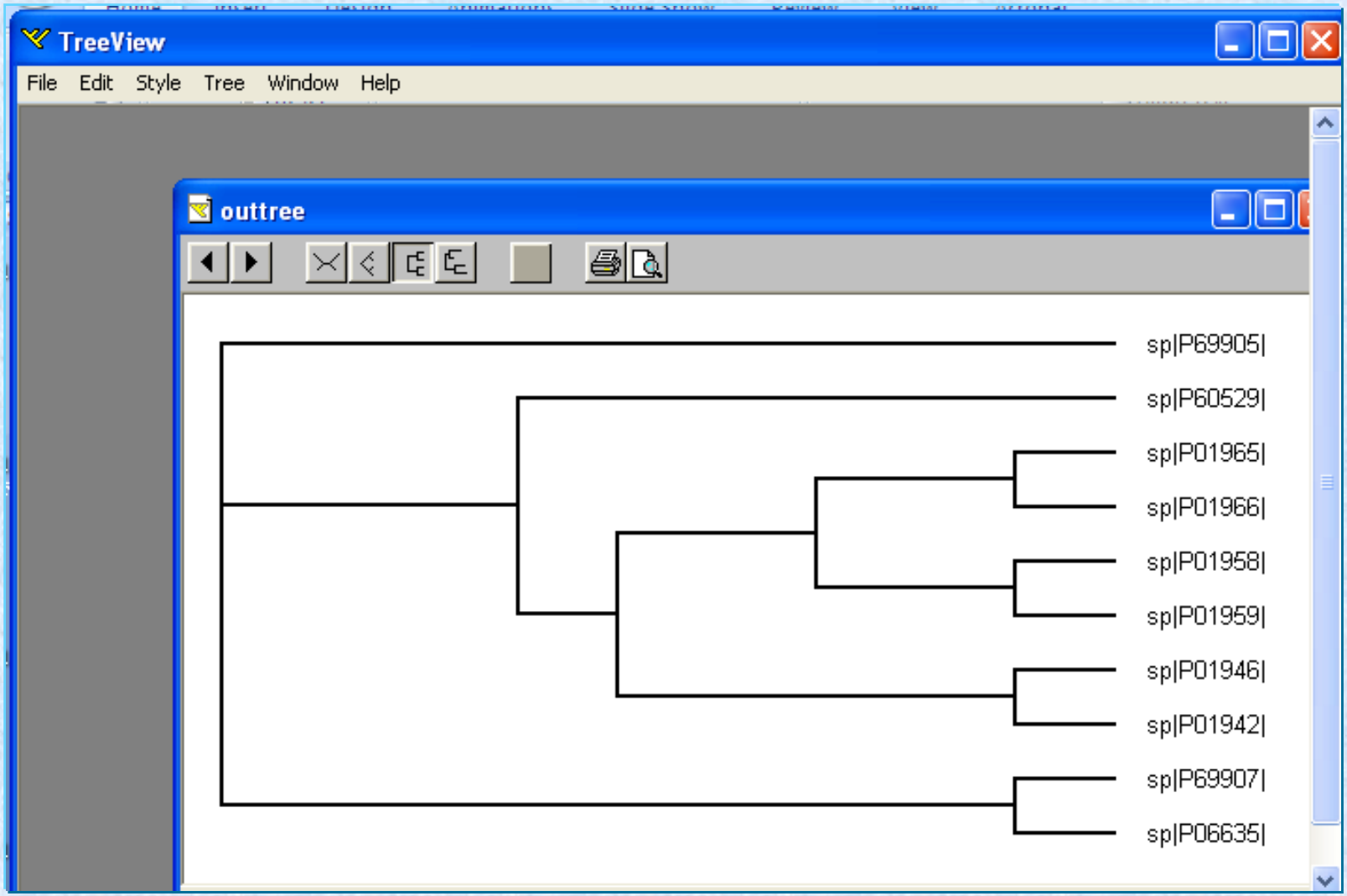
```
C:\Documents and Settings\Admin\Desktop\phylip
   4. sp¦P01959¦
   5. sp¦P60529¦
   6. sp¦P01946¦
   7. sp¦P69907¦
   8. sp¦P01958¦
   9. sp¦P01965¦
  10. sp¦P01942¦

Adding species:
   1. sp¦P01958¦
   2. sp¦P60529¦
   3. sp¦P01966¦
   4. sp¦P69907¦
   5. sp¦P69905¦
   6. sp¦P01959¦
   7. sp¦P06635¦
   8. sp¦P01965¦
   9. sp¦P01946¦
  10. sp¦P01942¦

Output written to file "work1a"

Tree also written onto file "outtree"

Done.

Press enter to quit.
```

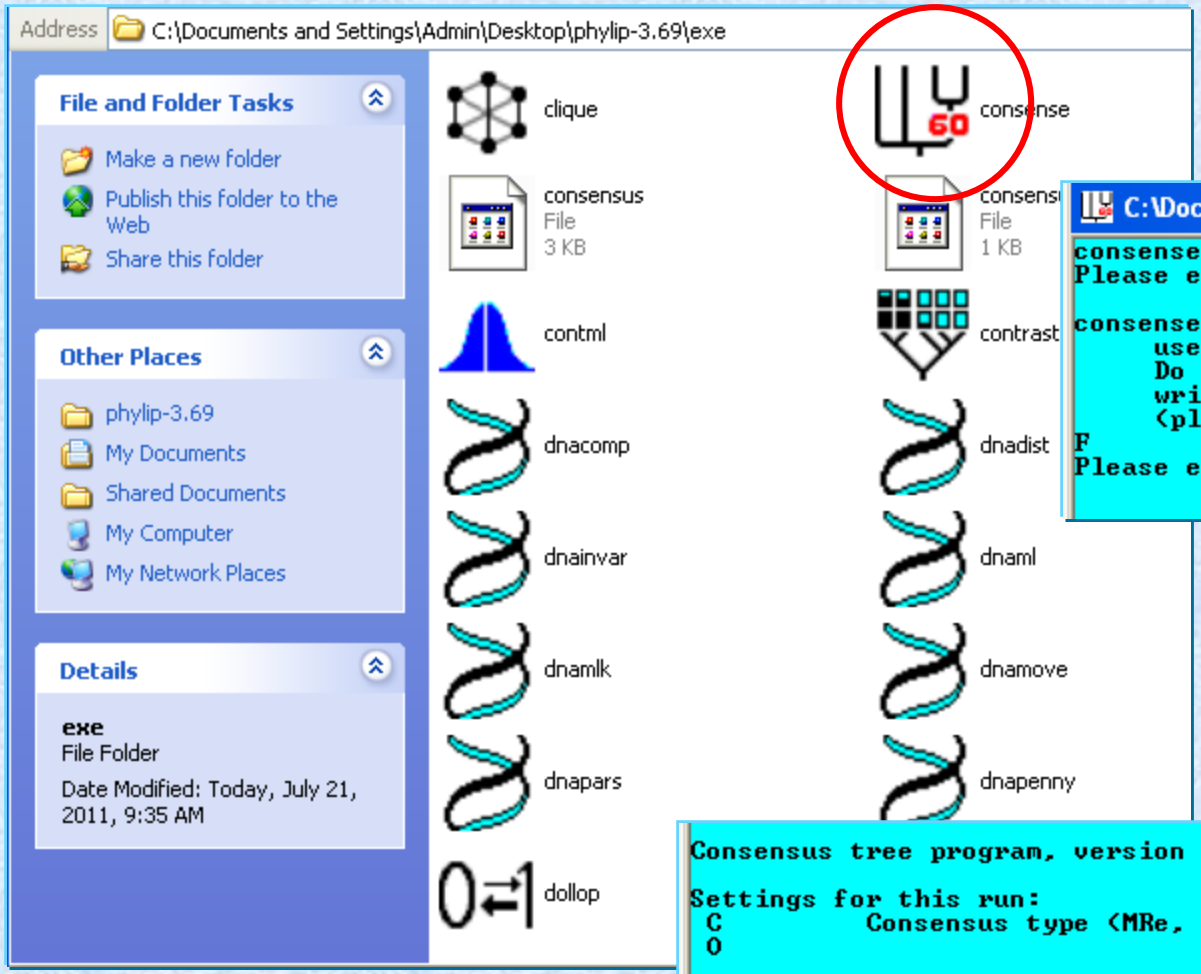



Admin
Internet
Internet Explorer
E-mail
Microsoft Office Outl
Microsoft Office Wor
Google Chrome
Adobe Acrobat 9 Pro
Microsoft Office Pow 2007
Notepad
WordPad
All Programs
start
Startup
VideoLAN
Internet Explorer
MSN
Outlook Express
Remote Assistance
Windows Media Player
Windows Messenger
Windows Movie Maker
Mozilla Firefox 4.0 Beta 8
ntlmaps
Acrobat Distiller 9
Acrobat.com
Adobe Acrobat 9 Pro
Adobe LiveCycle Designer ES 8.2
Adobe Reader 9
WinRAR
SSH Secure Shell
Google Chrome
PyMOL
mEmboss
Modeller9.9
TreeView


TreeView
File Edit Style Tree Window Help
New Ctrl+N
Open... Ctrl+O
Close Ctrl+W
Save Ctrl+S
Save as...
Save as graphic...
Import
Page setup...
Print preview
Print... Ctrl+P
1 c:\...\consensustree
Exit Ctrl+Q


Open
Look in: exe
drawtree
factor
fitch
font1
font2
font3
font4
font5
font6
gendist
hemoglobin
hemoglobin2
hemoglobin3
hemoglobintree
kitsch
mix
move
neighbor
outfile
outtree
pars
penny
proml
promlk
File name: outtree
Open
Files of type: All files (*.*)
Cancel


TreeView
File
Edit
Style
Tree
Window
Help
outtree
sp|P69905|
sp|P60529|
sp|P01965|
sp|P01966|
sp|P01958|
sp|P01959|
sp|P01946|
sp|P01942|
sp|P69907|
sp|P06635|

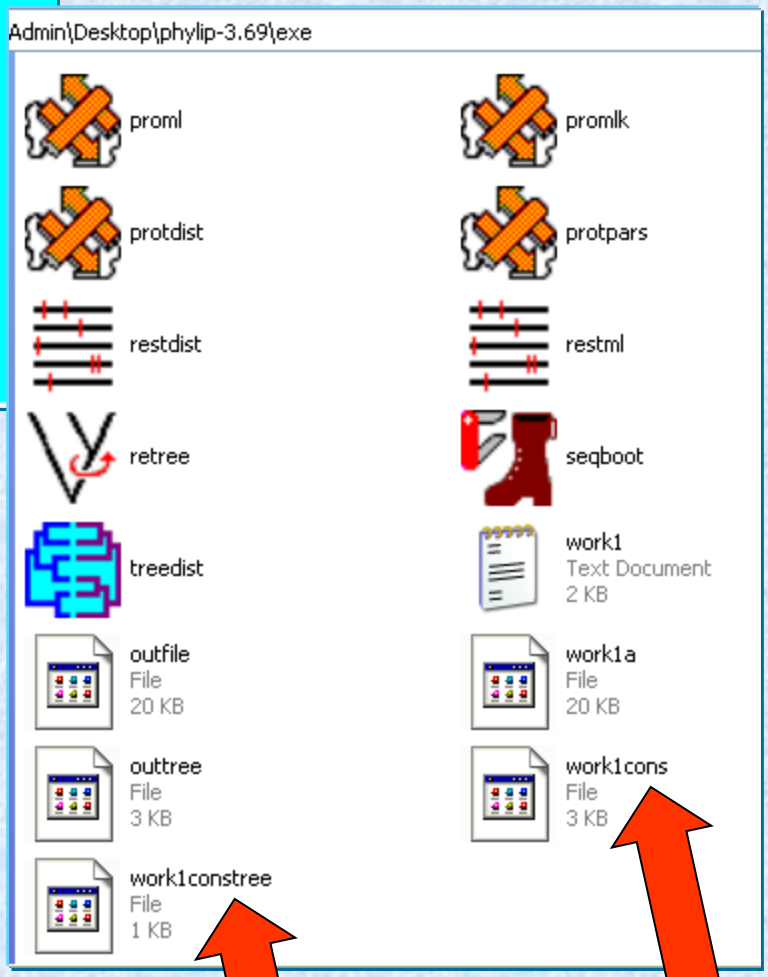


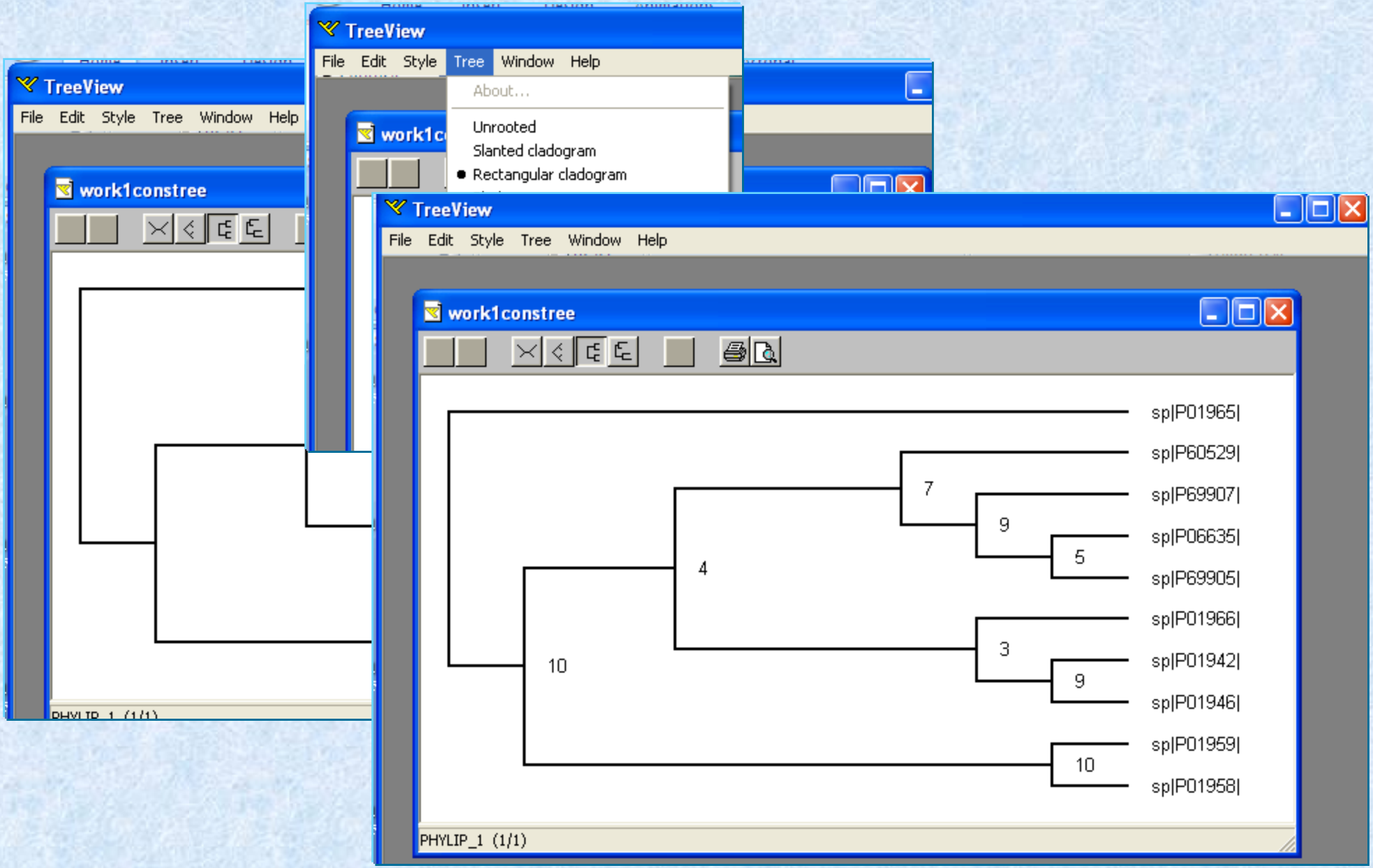

```
C:\Documents and Settings\Admin\Desktop\phylip-3.69\exe\consei
consense.exe: can't find input tree file "intree"
Please enter a new file name> outtree

consense.exe: the file "outfile" that you wanted to
     use as output file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
F
Please enter a new file name> work1cons
```

```
Consensus tree program, version 3.69

Settings for this run:
 C          Consensus type (MRe, strict, MR, Ml):  Majority rule (extended)
 O                                Outgroup root:  No, use as outgroup species  1

 R                 Trees to be treated as Rooted:  No
 T             Terminal type (IBM PC, ANSI, none):  IBM PC
 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)
y_
```

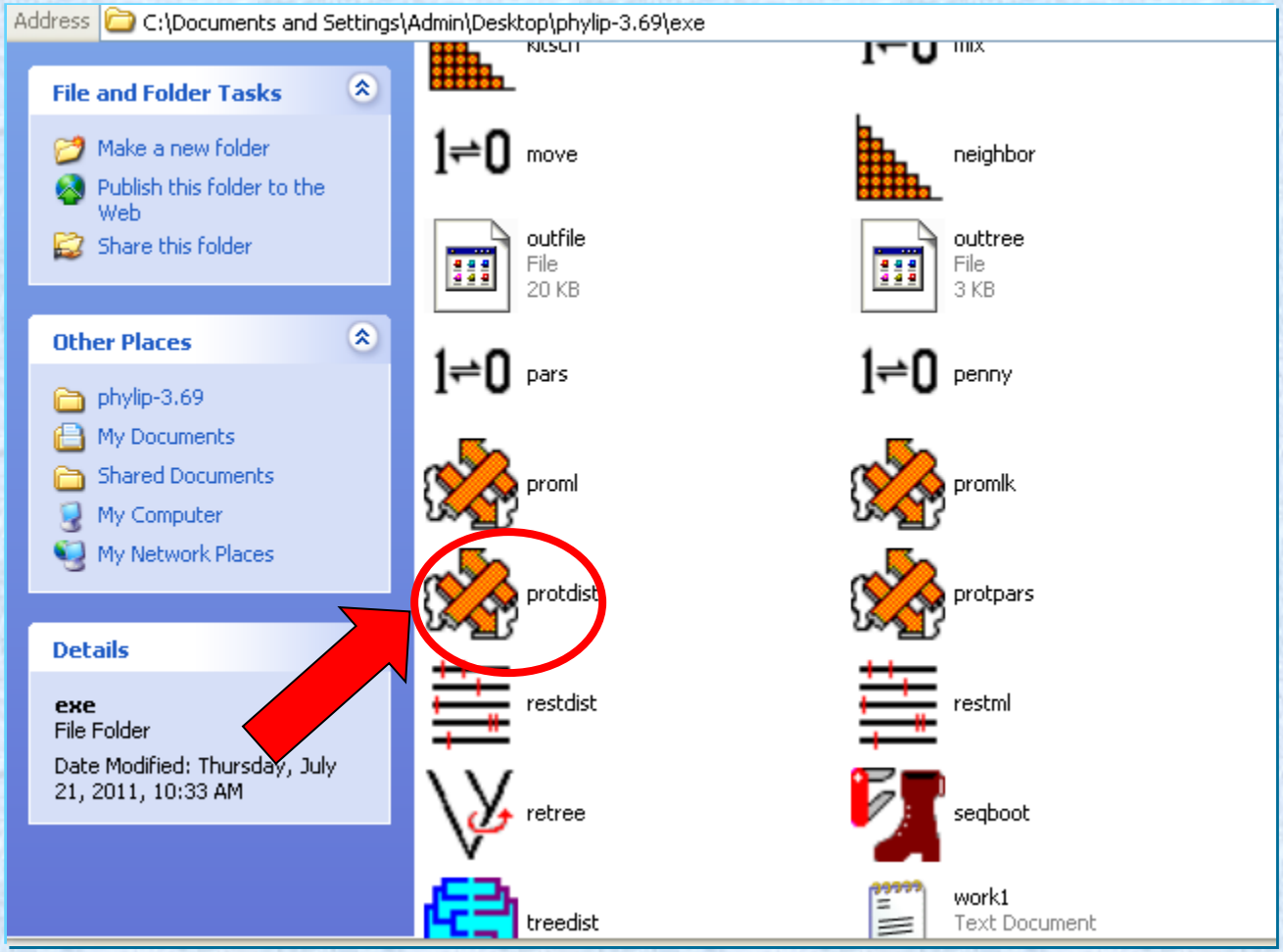
```
Consensus tree program, version 3.69

Settings for this run:
 C         Consensus type (MRe, strict, MR, Ml):  Majority rule (extended)
 O                              Outgroup root:  No, use as outgroup species  1

 R               Trees to be treated as Rooted:  No
 T           Terminal type (IBM PC, ANSI, none):  IBM PC
 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)
y

consense.exe: the file "outtree" that you wanted to
     use as output tree file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
f
Please enter a new file name> work1constree_
```

```
Consensus tree written to file "work1constree"

Output written to file "work1cons"

Done.

Press enter to quit.
```




TreeView
File Edit Style Tree Window Help
About...
Unrooted
Slanted cladogram
Rectangular cladogram
work1constree
sp|P01965|
sp|P60529|
sp|P69907|
sp|P06635|
sp|P69905|
sp|P01966|
sp|P01942|
sp|P01946|
sp|P01959|
sp|P01958|
7
9
5
4
3
9
10
10
PHYLIP_1 (1/1)

NJ and UPGMA methods



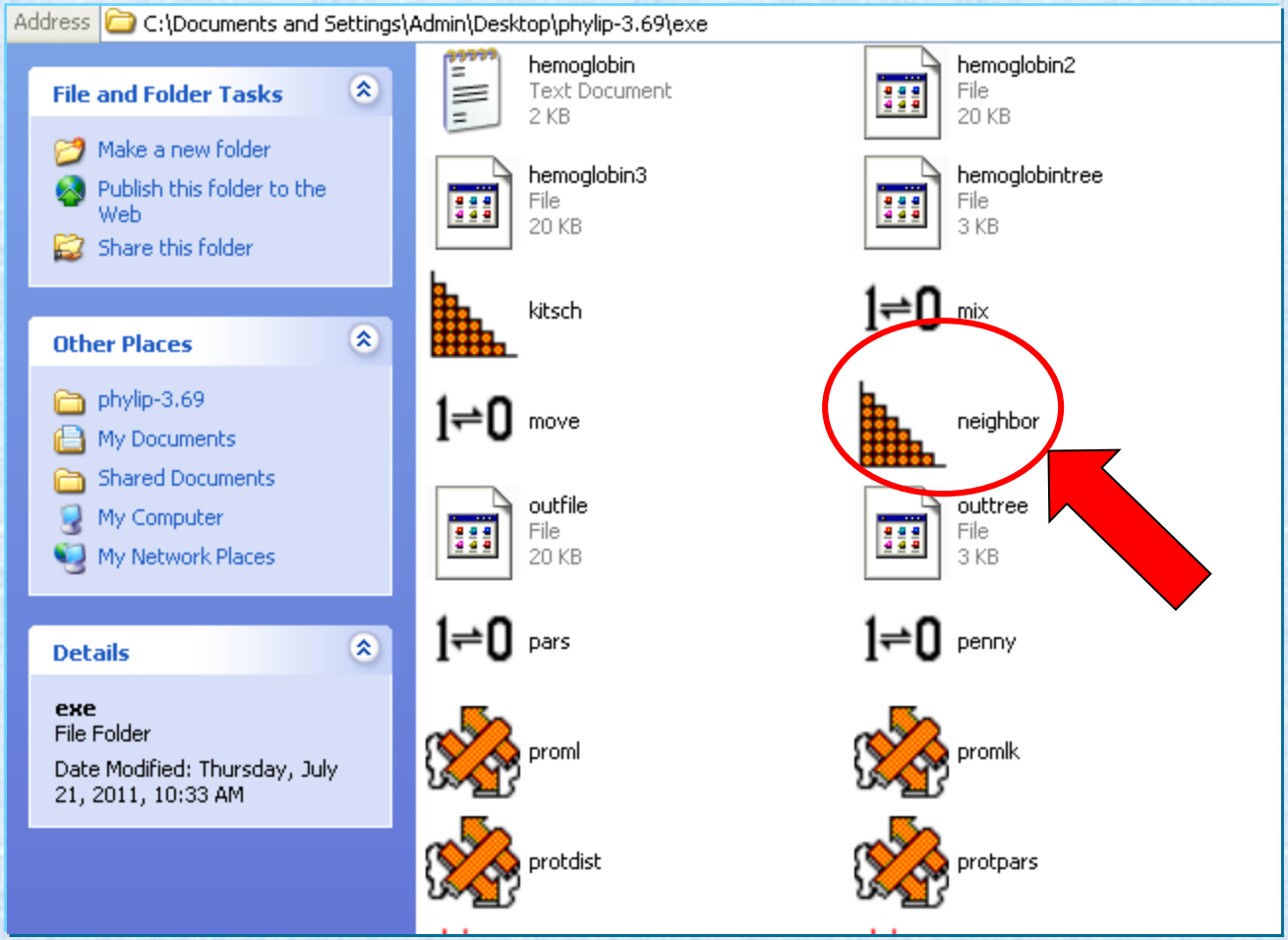
```
protdist.exe: can't find input file "infile"
Please enter a new file name> outfile

protdist.exe: the file "outfile" that you wanted to
     use as output file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
f
Please enter a new file name> work1-protdist
```

```
Protein distance algorithm, version 3.69

Settings for this run:
  P  Use JTT, PMB, PAM, Kimura, categories model?  Jones-Taylor-Thornton matrix
  G  Gamma distribution of rates among positions?  No
  C           One category of substitution rates?  Yes
  W                    Use weights for positions?  No
  M                   Analyze multiple data sets?  No
  I                  Input sequences interleaved?  Yes
  0                Terminal type (IBM PC, ANSI)?  IBM PC
  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)
m
Multiple data sets or multiple weights? (type D or W)
d
How many data sets?
10
```

```
Protein distance algorithm, version 3.69

Settings for this run:
  P  Use JTT, PMB, PAM, Kimura, categories model?  Dayhoff PAM matrix
  G  Gamma distribution of rates among positions?  No
  C           One category of substitution rates?  Yes
  W                    Use weights for positions?  No
  M                   Analyze multiple data sets?  Yes, 10 data sets
  I                  Input sequences interleaved?  Yes
  0                Terminal type (IBM PC, ANSI)?  IBM PC
  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)
```

work1-protdist

```
sp|Q9WVA2|  0.000000  0.954472  1.211022  1.534523  2.384227
2.197177
  2.466581  2.303897  2.378322  2.245130  2.098663  2.066428
1.923666
  1.946335
sp|Q9Y5J9|  0.954472  0.000000  1.114781  1.699316  2.007548
2.014938
  2.236423  1.873154  2.401207  2.523799  2.217026  2.237688
1.818816
  1.896587
sp|Q09783|  1.211022  1.114781  0.000000  1.303653  2.771608
2.657542
  2.794046  2.191316  2.722125  2.545494  2.061901  2.150728
2.356218
  2.009404
sp|Q75DU7|  1.534523  1.699316  1.303653  0.000000  2.572591
2.581259
  2.958788  2.454755  2.303727  2.134035  2.242738  1.951797
1.613806
  2.068258
sp|Q75F72|  2.384227  2.007548  2.771608  2.572591  0.000000
0.263512
  0.532918  0.949154  1.037593  1.164635  1.410843  1.465146
1.892952
  2.173354
sp|Q6CJX3|  2.197177  2.014938  2.657542  2.581259  0.263512
0.000000
  0.454437  0.916289  1.004302  1.030748  1.405926  1.425100
1.939355
```

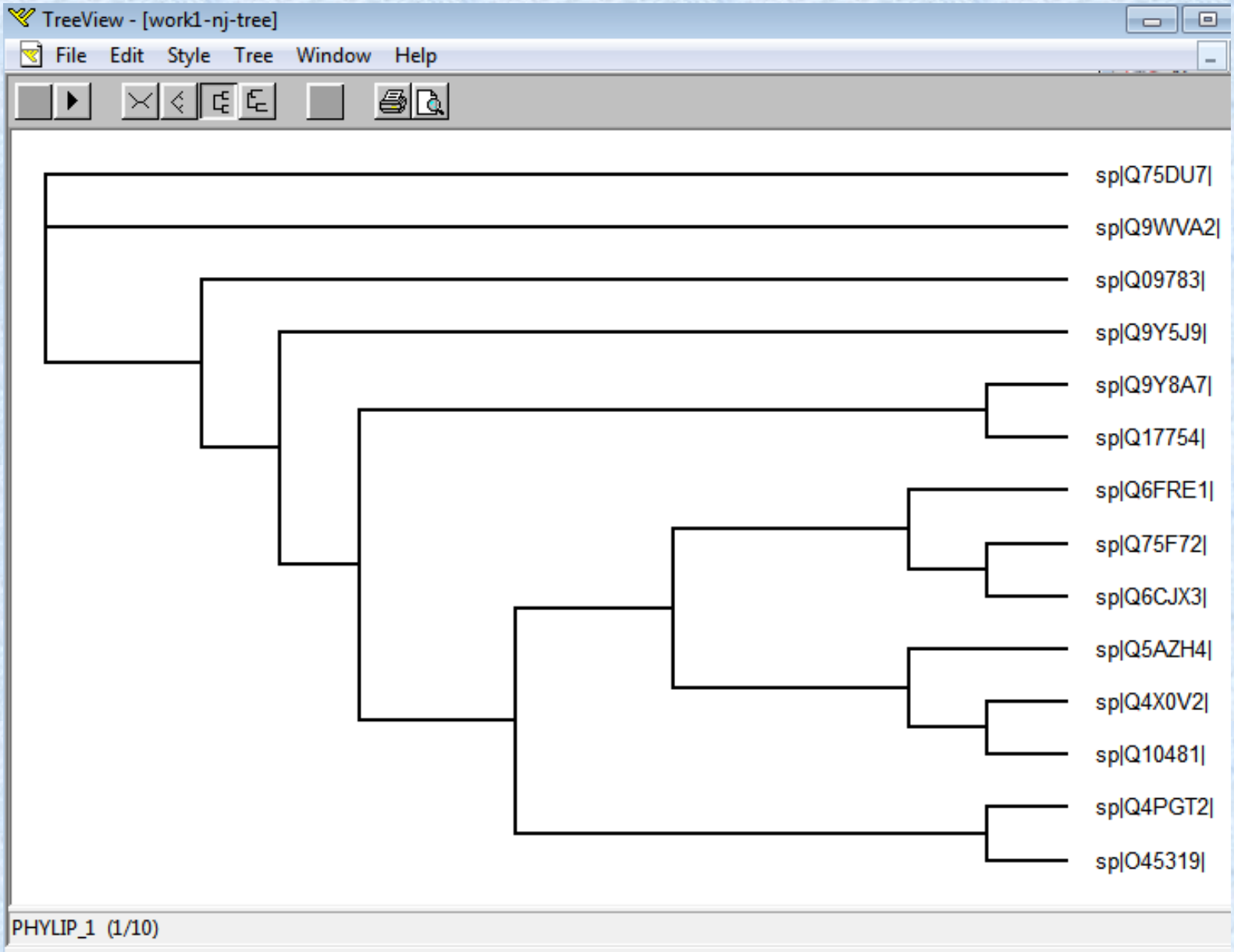

Address
C:\Documents and Settings\Admin\Desktop\phylip-3.69\exe
File and Folder Tasks
Make a new folder
Publish this folder to the Web
Share this folder
Other Places
phylip-3.69
My Documents
Shared Documents
My Computer
My Network Places
Details
exe
File Folder
Date Modified: Thursday, July 21, 2011, 10:33 AM
hemoglobin
Text Document
2 KB
hemoglobin2
File
20 KB
hemoglobin3
File
20 KB
hemoglobintree
File
3 KB
kitsch
mix
move
neighbor
outfile
File
20 KB
outtree
File
3 KB
pars
penny
proml
promlk
protdist
protpars

```
neighbor.exe: can't find input file "infile"
Please enter a new file name> work1-protdist

neighbor.exe: the file "outfile" that you wanted to
     use as output file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
F
Please enter a new file name> work1-nj
```

```
Neighbor-Joining/UPGMA method version 3.69

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, ANSI, none)?  IBM PC
  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


  Y to accept these or type the letter for one to change
m
How many data sets?
10

Random number seed (must be odd)?
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?  Yes (random number seed =       5)
  M           Analyze multiple data sets?  Yes, 10 sets
  0   Terminal type (IBM PC, ANSI, none)?  IBM PC
  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


  Y to accept these or type the letter for one to change
y

neighbor.exe: the file "outtree" that you wanted to
     use as output tree file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
f
Please enter a new file name> work1-nj-tree
```

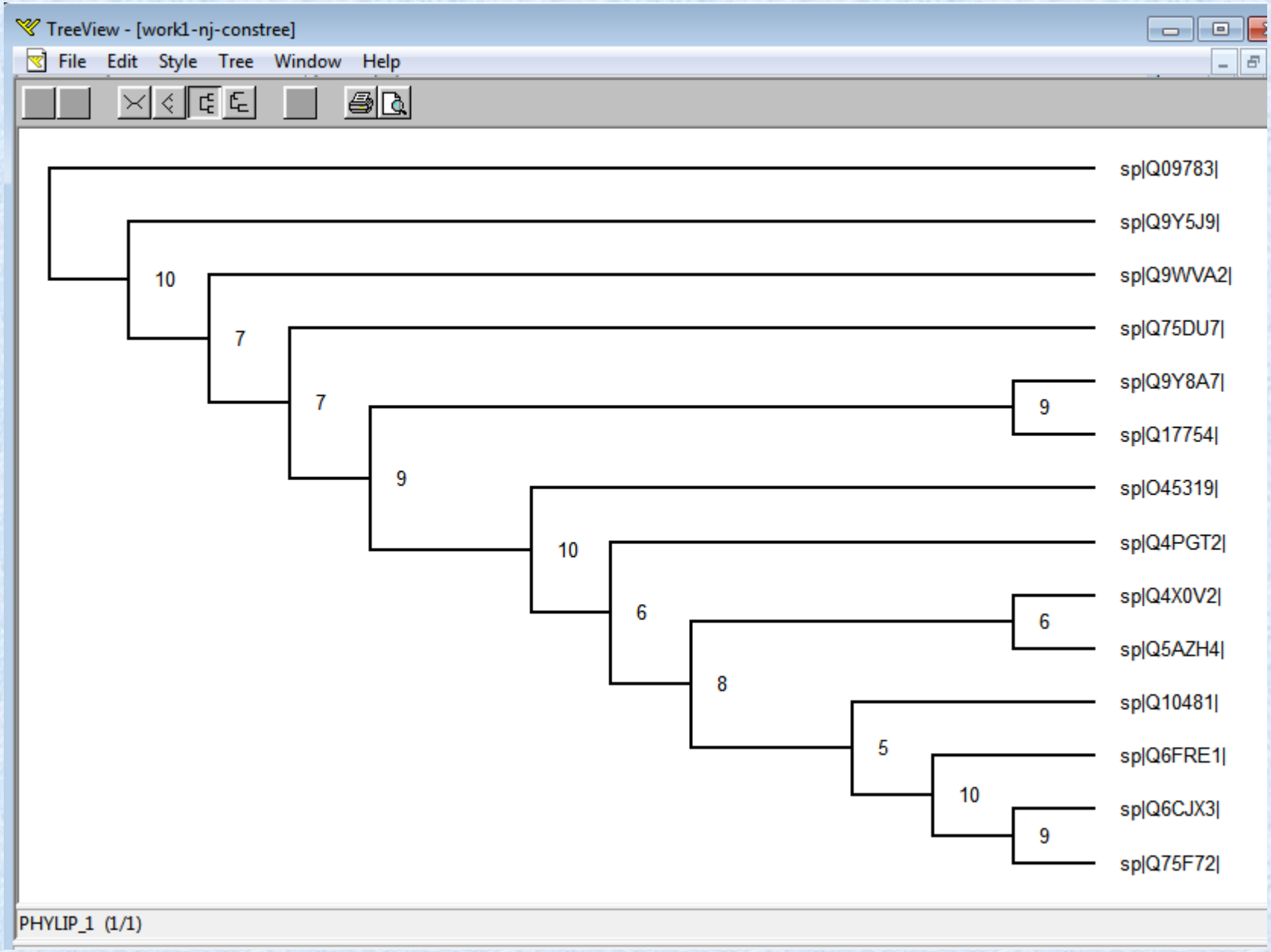
```
((sp|Q09783|:0.61703,(sp|Q9Y5J9|:0.71119,((sp|Q9Y8A7|:0.96547,
sp|Q17754|:0.59919):0.21205,(((sp|Q6FRE1|:0.32455,(sp|Q75F72
|:0.10986,
sp|Q6CJX3|:0.11552):0.07268):0.29205,((sp|Q4X0V2|:0.40667,
sp|Q10481|:0.30864):0.10478,sp|Q5AZH4
|:0.45459):0.12561):0.17539,
(sp|Q4PGT2|:0.67722,sp|O45319
|:0.67765):0.02625):0.50607):0.47300):0.27258):0.09925,
sp|Q75DU7|:0.56167,sp|Q9WVA2|:0.64156);
((sp|Q09783|:0.68210,sp|Q9Y5J9|:0.38881):0.03306,(((sp|Q9Y8A7
|:0.64578,
sp|Q17754|:0.49015):0.06214,((sp|O45319|:0.50317,sp|Q4PGT2
|:0.42263):0.12931,
(sp|Q5AZH4|:0.36428,(sp|Q4X0V2|:0.39579,(((sp|Q75F72|:0.12829,
sp|Q6CJX3|:0.06919):0.12728,sp|Q6FRE1|:0.29582):0.35209,
sp|Q10481|:0.38144):0.03146):0.07114):0.13883):0.70546):0.52085,
sp|Q75DU7|:0.77374):0.34769,sp|Q9WVA2|:0.46575);
((sp|Q9Y5J9|:0.30392,sp|Q09783|:0.67455):0.07237,((sp|Q17754
|:0.56886,
((sp|O45319|:0.52738,(sp|Q4PGT2|:0.49822,(sp|Q10481|:0.35123,
(((sp|Q6CJX3|:0.08839,sp|Q75F72|:0.12819):0.10053,sp|Q6FRE1
|:0.26834):0.28216,
(sp|Q5AZH4|:0.33943,sp|Q4X0V2
|:0.30149):0.12912):0.02663):0.11009):0.03799):0.40163,
sp|Q75DU7|:0.96568):0.13386):0.11587,sp|Q9Y8A7
|:0.62374):0.33497,sp|Q9WVA2|:0.37622);
((sp|Q75DU7|:0.81103,((sp|O45319|:0.60856,((((sp|Q6FRE1
|:0.36825,
sp|Q6CJX3|:0.11354):0.02000,sp|Q75F72|:0.15488):0.34593,
sp|Q10481|:0.29363):0.02586,((sp|Q4PGT2|:0.56535,sp|Q5AZH4
|:0.38322):0.10438,
sp|Q4X0V2|:0.40429):0.07203):0.20545):0.71213,(sp|Q9Y8A7
```


TreeView - [work1-nj-tree]
File Edit Style Tree Window Help
sp|Q75DU7|
sp|Q9WVA2|
sp|Q09783|
sp|Q9Y5J9|
sp|Q9Y8A7|
sp|Q17754|
sp|Q6FRE1|
sp|Q75F72|
sp|Q6CJX3|
sp|Q5AZH4|
sp|Q4X0V2|
sp|Q10481|
sp|Q4PGT2|
sp|O45319|
PHYLIP_1 (1/10)

```
consense.exe: can't find input tree file "intree"
Please enter a new file name> work1-nj-tree

consense.exe: the file "outfile" that you wanted to
     use as output file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
f
Please enter a new file name> work1-nj-cons
```

```
Consensus tree program, version 3.695

Settings for this run:
 C         Consensus type (MRe, strict, MR, Ml):  Majority rule (extended)
 O                                Outgroup root:  No, use as outgroup species

 R               Trees to be treated as Rooted:  No
 T           Terminal type (IBM PC, ANSI, none):  IBM PC
 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)
y

consense.exe: the file "outtree" that you wanted to
     use as output tree file already exists.
     Do you want to Replace it, Append to it,
     write to a new File, or Quit?
     (please type R, A, F, or Q)
f
Please enter a new file name> work1-nj-constree
```

```
((((((sp|Q9Y8A7|:10.0,sp|Q17754|:10.0):9.00,((((sp|Q4X0V2
|:10.0,sp|Q5AZH4|:10.0):6.00,
(((sp|Q6CJX3|:10.0,sp|Q75F72|:10.0):9.00,sp|Q6FRE1
|:10.0):10.0,sp|Q10481|:10.0):5.00):8.00,
sp|Q4PGT2|:10.0):6.00,sp|O45319|:10.0):10.0):9.00,sp|Q75DU7
|:10.0):7.00,sp|Q9WVA2|:10.0):7.00,
sp|Q9Y5J9|:10.0):10.0,sp|Q09783|:10.0);
```


TreeView - [work1-nj-constree]
File Edit Style Tree Window Help
sp|Q09783|
sp|Q9Y5J9|
sp|Q9WVA2|
sp|Q75DU7|
sp|Q9Y8A7|
sp|Q17754|
sp|O45319|
sp|Q4PGT2|
sp|Q4X0V2|
sp|Q5AZH4|
sp|Q10481|
sp|Q6FRE1|
sp|Q6CJX3|
sp|Q75F72|
10
7
7
9
9
10
6
6
8
5
10
9
PHYLIP_1 (1/1)