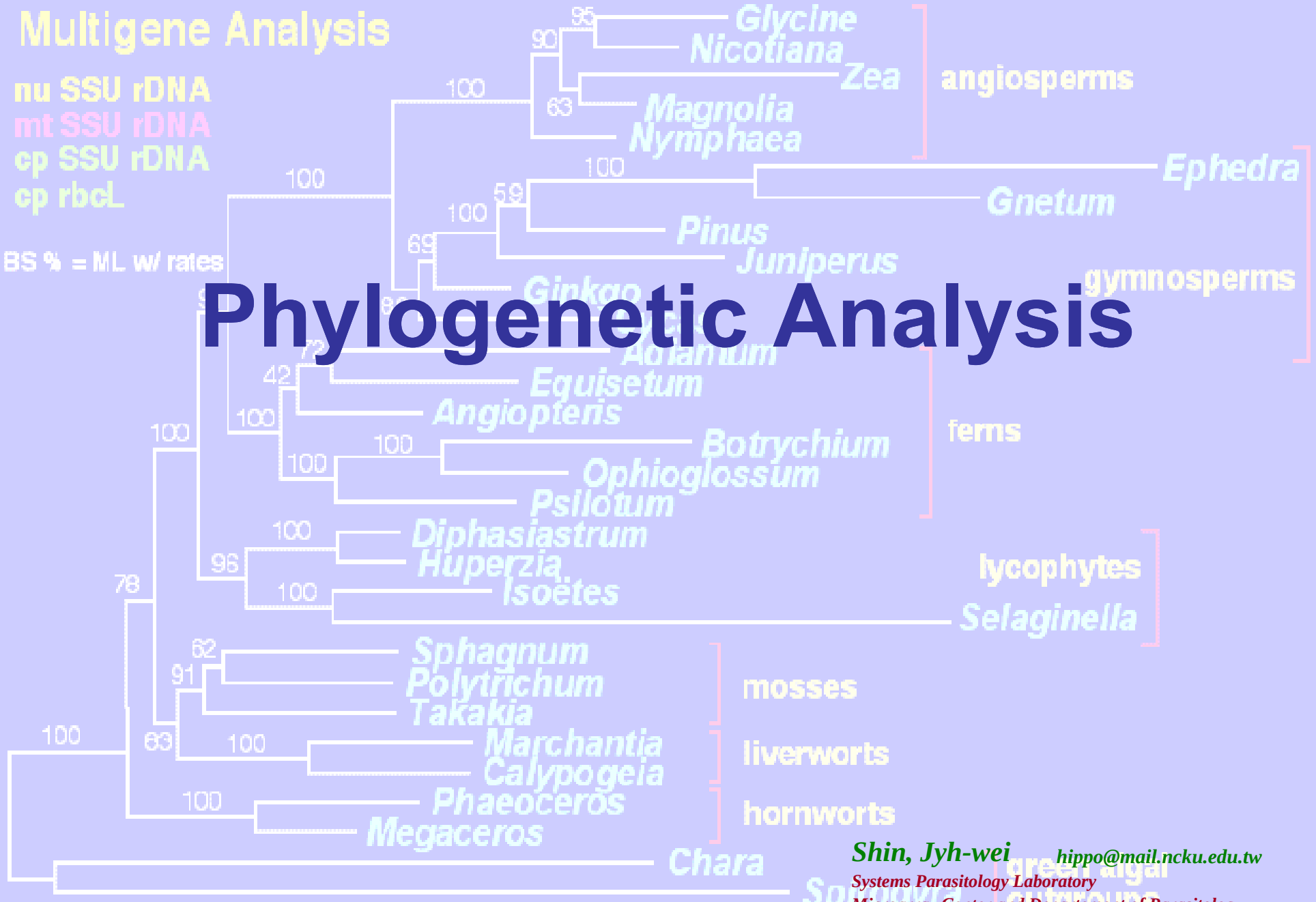
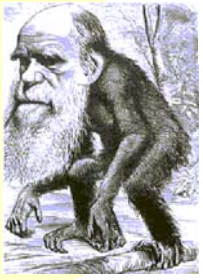


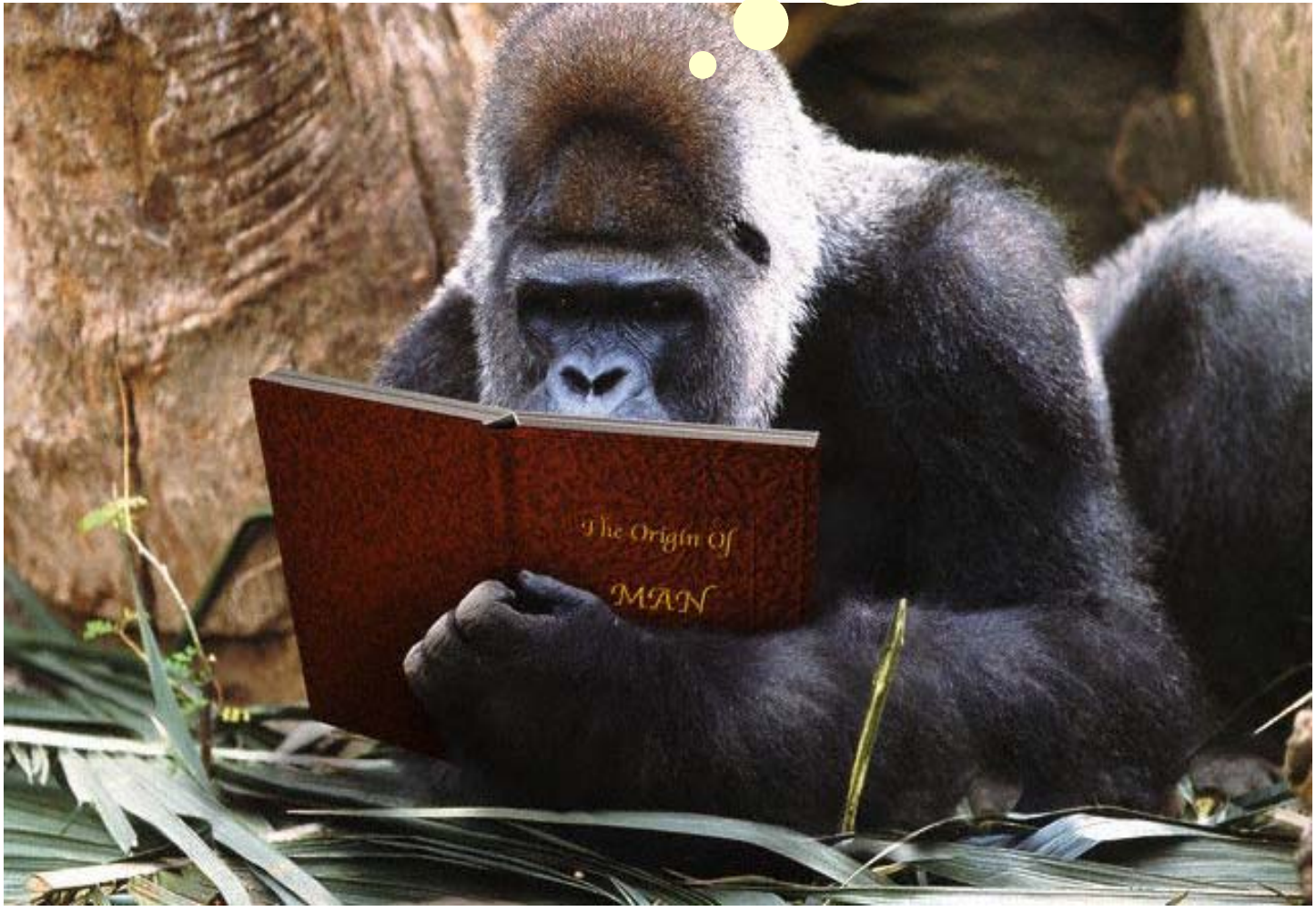
Phylogenetic Analysis

Shin, Jyh-wei *hippo@mail.ncku.edu.tw*
Systems Parasitology Laboratory
Microarray Center and Departement of Parasitology
College of Medicine, National Chung Kung UNiversity

Nature Selection

Can we doubt … that individuals having any advantage, however slight, over others, would have the best chance of surviving and proceeding their kind? On the other hand, we may feel sure that any variation in the least degree injurious would be rigidly destroyed variations, I call Nature Selection.





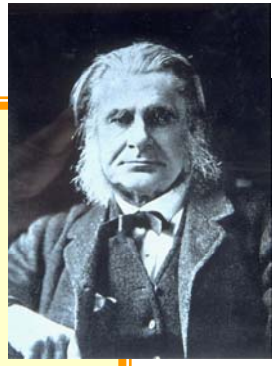
Phylogenetic systematics

The identification and analysis of homologies is central to phylogenetic systematics

- Sees homology as evidence of common ancestry
- Uses tree diagrams to portray relationships based upon recency of common ancestry
- Monophyletic groups (clades) - contain species which are more closely related to each other than to any outside of the group



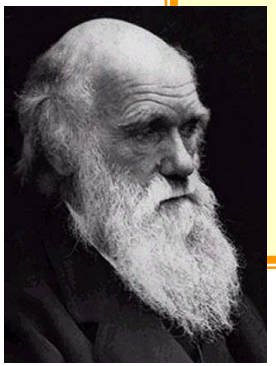
Darwin's letter to Thomas Huxley 1857

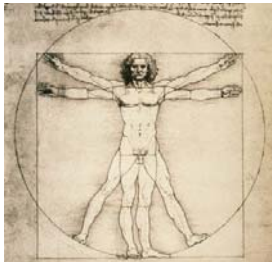


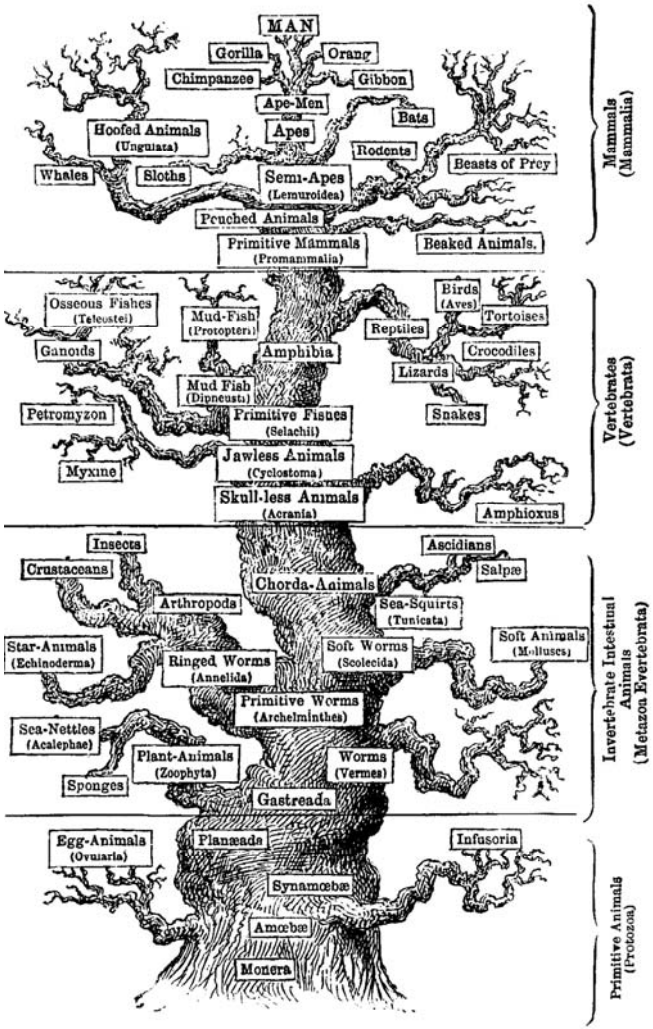
Dear Thomas,

The time will come I believe, though I shall not live to see it, when we shall have fairly true genealogical (phylogenetic) trees of each great kingdom of nature.

Charles Darwin





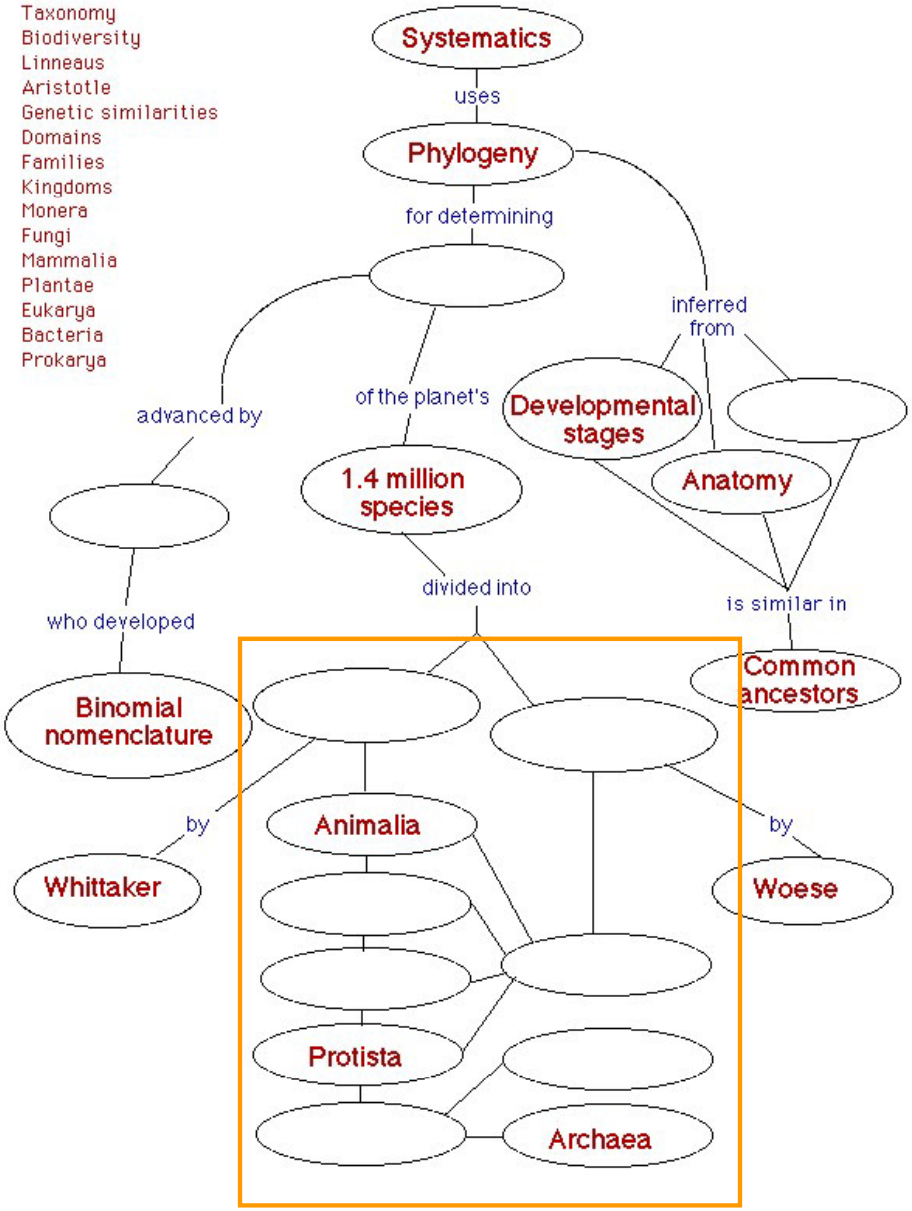


Haeckel's pedigree of man

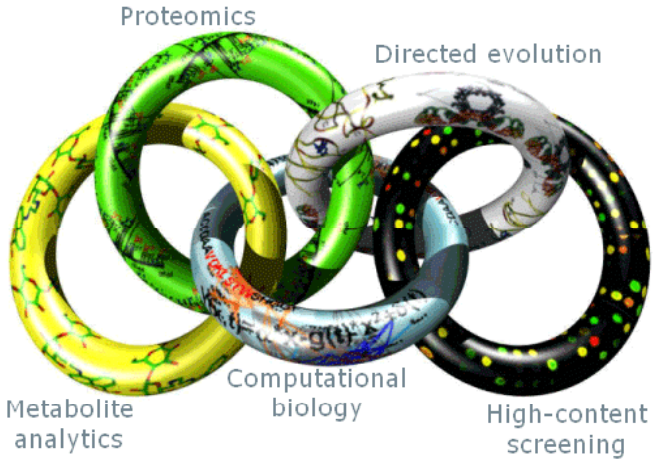
Systematics: Field of biology that deals with the diversity of life. Systematics is usually divided into the two areas of phylogenetics and taxonomy

Phylogenetics: Field of biology that studies the evolutionary relationships between organisms. It includes the discovery of these relationships, and the study of the causes behind this pattern

Taxonomy: The science of naming and classifying organisms



SYSTEMS BIOLOGY



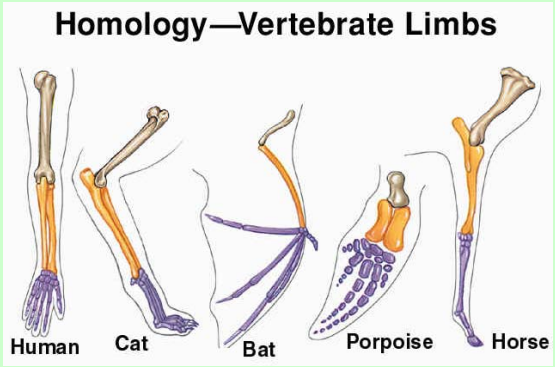
http://www.biology.lsu.edu/introbio/tutorial/Concept-maps/1002/systematics-map.html
http://www.cmdr.ubc.ca/pathogenomics/terminology.html

Homology is...

The relationship of any two characters that have descended from a common ancestor. This term can apply to a morphological structure, a chromosome or an individual gene or DNA segment.

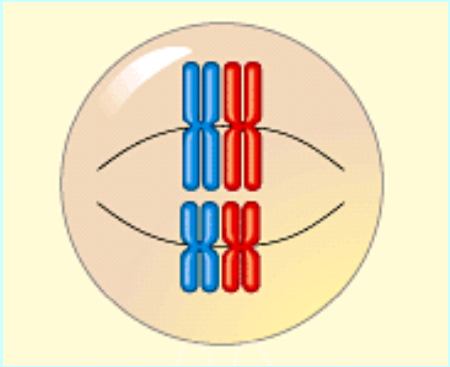
Homologous structure

Characters in different species which were inherited from a common ancestor and thus share a similar ontogenetic pattern.



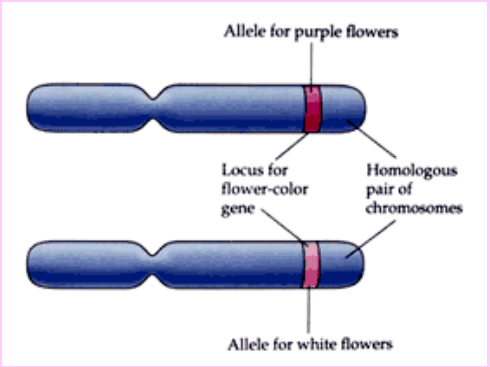
Homologous chromosome

One part of two genetically different chromosomes. Each homologous chromosome is inherited from a different parent, and contains information about the same gene sequence.



Homologous gene

Molecular investigations by developmental biologists have revealed striking similarities between the structure of genes (The hereditary determinant of a specified characteristic of an individual; specific sequences of nucleotides in DNA.) regulating ontogenetic phenomena in diverse organisms.



Homology is... They said that ……

Homologue:

the same organ under every variety of form and function (true or essential correspondence)

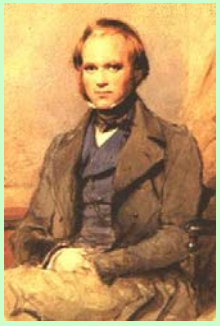
Analogy:

superficial or misleading similarity



Richard Owen 1843

"The natural system is based upon descent with modification .. the characters that naturalists consider as showing true affinity (i.e. homologies) are those which have been inherited from a common parent, and, in so far as all true classification is genealogical; that community of descent is the common bond that naturalists have been seeking"



Charles Darwin, Origin of species 1859 p. 413

Cladistic vs. Phenetic

Within the field of taxonomy there are two different methods and philosophies of building phylogenetic trees: cladistic and phenetic.

Cladistic methods rely on assumptions about ancestral relationships as well as on current data.

- For character data about the physical traits of organisms (such as morphology of organs etc.) and for deeper levels of taxonomy, the cladistic approach is almost certainly superior.
- Cladistic methods are often difficult to implement with molecular data because all of the assumptions are generally not satisfied.

Phenetic methods construct trees (phenograms) by considering the current states of characters without regard to the evolutionary history that brought the species to their current phenotypes.

- Computer algorithms based on the phenetic model rely on Distance Methods to build of trees from sequence data.
- Phenetic methods count each base of sequence difference equally, so a single event that creates a large change in sequence (insertion/deletion or recombination) will move two sequences far apart on the final tree.
- Phenetic approaches generally lead to faster algorithms and they often have nicer statistical properties for molecular data.
- The phenetic approach is popular with molecular evolutionists because it relies heavily on objective character data (such as sequences) and it requires relatively few assumptions.

Cladograms and Phylograms

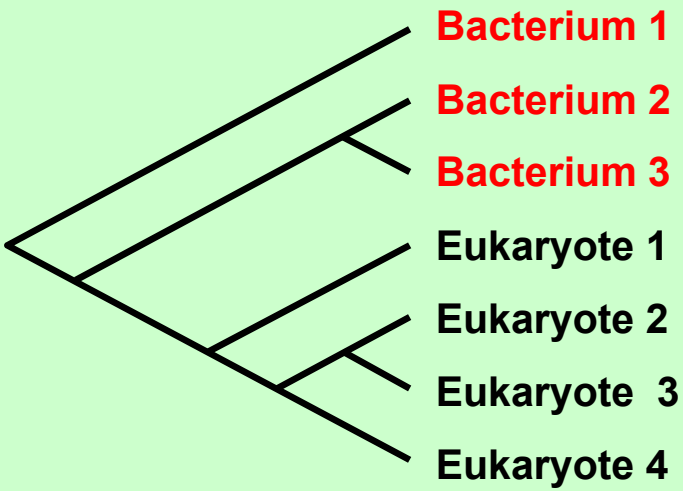
Cladograms

show branching order

and branch lengths are meaningless

分支圖 (cladograms)

表示現存與化石物種彼此的關係，
並非祖先或子嗣的關係。



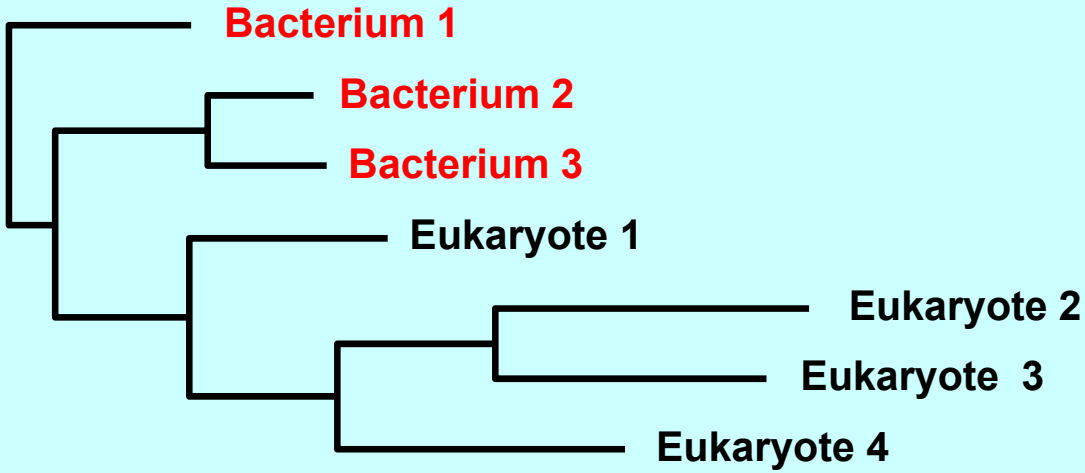
Phylograms

show branch order

and branch lengths

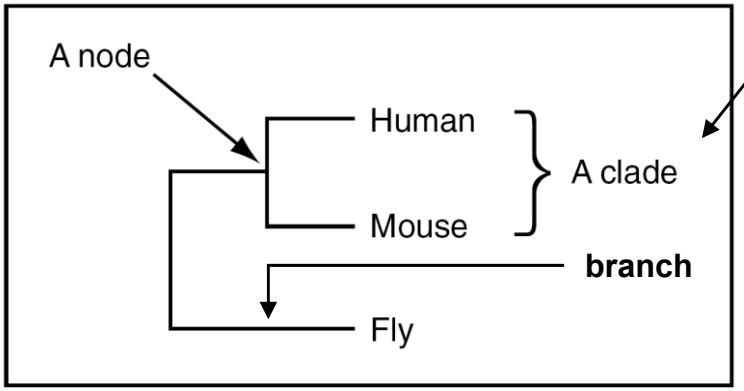
系統發生圖 (phylograms)

描述一群有機體發生或進化順
序的拓撲結構。



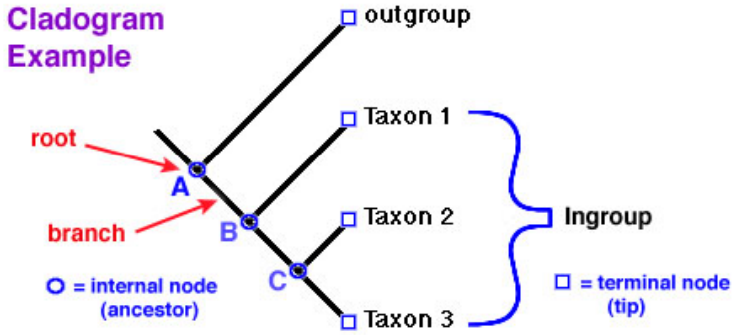
3 three basic assumptions in cladistics（遺傳分類學）

1. Any group of organisms is related by descent from a common ancestor.
2. There is a bifurcating pattern of cladogenesis. This assumption is controversial.
3. Change in characteristics occurs in lineages over time.

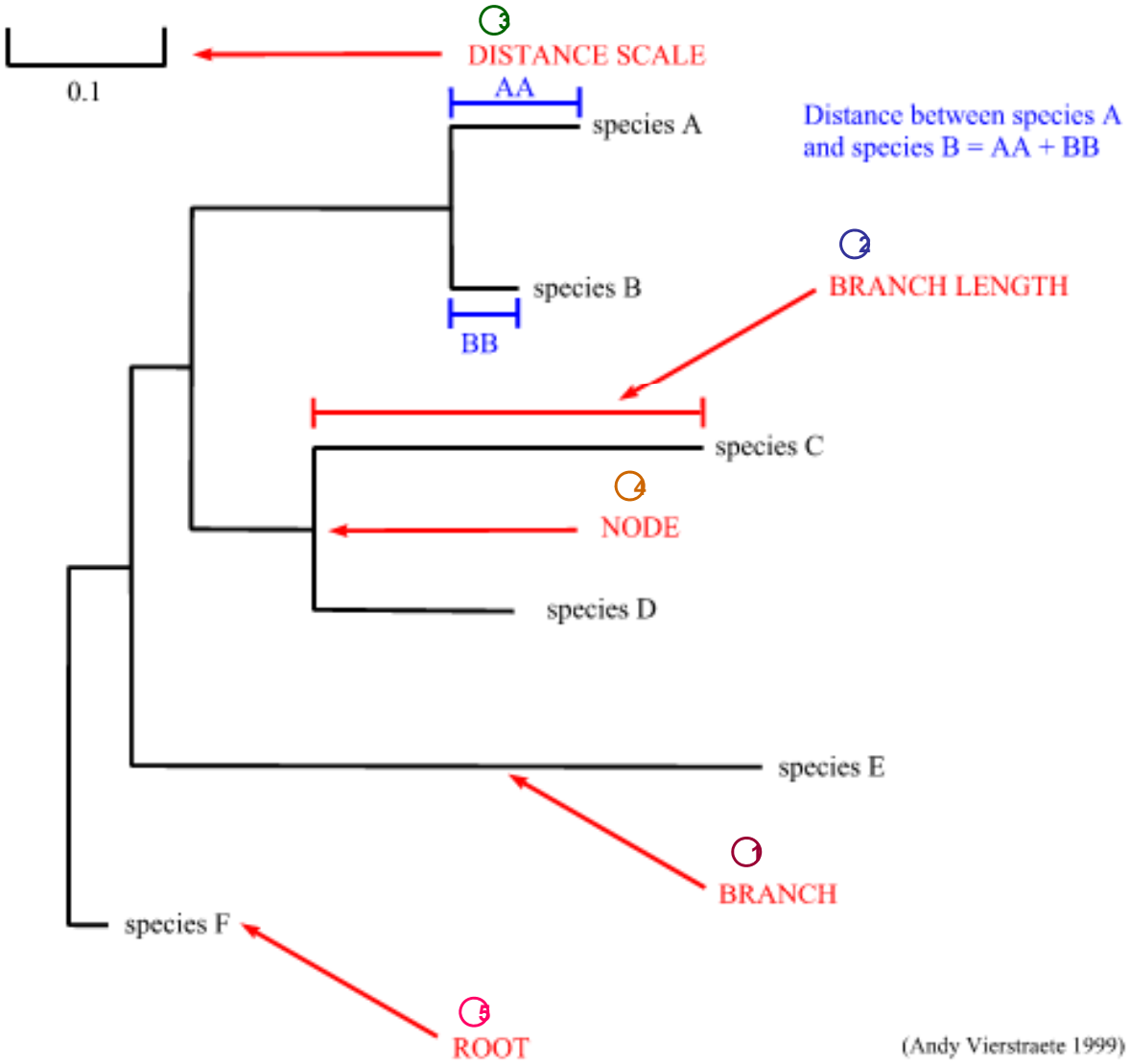


Clades are groups of organisms or genes that include the most recent common ancestor of all of its members and all of the descendants of that most recent common ancestor.
Clade is derived from the Greek word "klados," meaning branch or twig.

- **clade** 【群】 is a monophyletic taxon
- **taxon** 【分類群】 is any named group of organisms but not necessarily a clade
- **branch lengths** correspond to divergence
- **node** is a bifurcating branch point.



Tree Terminology



① **branch : defines the relationship between the taxa in terms of descent and ancestry**

② **branch length : often represents the number of changes that have occurred in that branch**

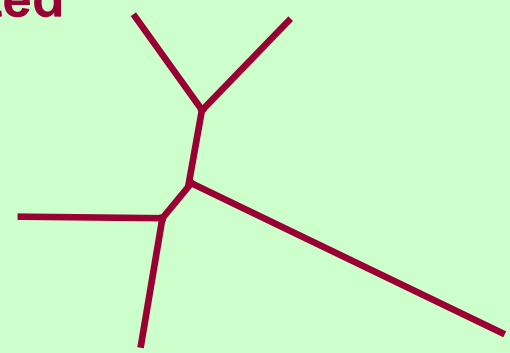
③ **distance scale :** scale which represents the number of differences between sequences (e.g. 0.1 means 10 % differences between two sequences)

④ **node : a node represents a taxonomic unit. This can be a taxon (an existing species) or an ancestor (unknown species : represents the ancestor of 2 or more species).**

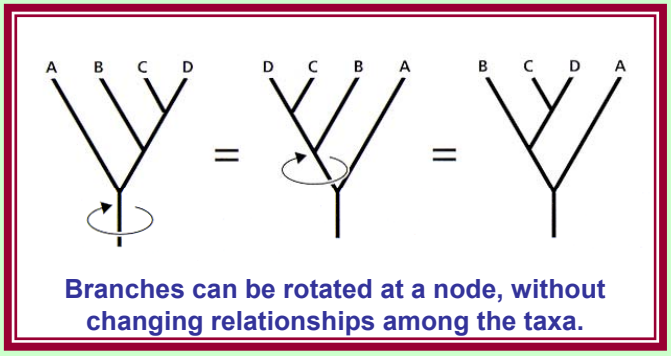
⑤ **root : is the common ancestor of all taxa**

Unrooted versus rooted phylogenies

unrooted

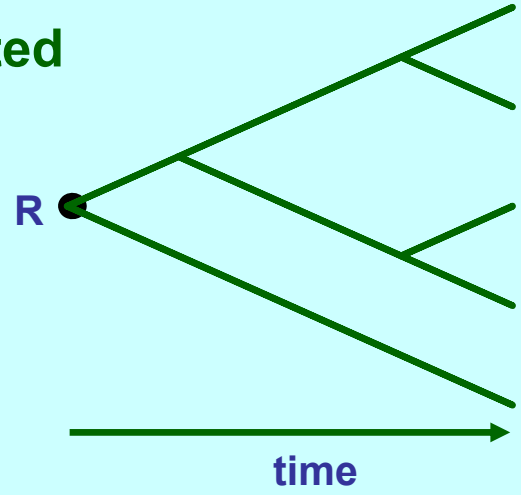


only specifies relationships not the evolutionary path



Branches can be rotated at a node, without changing relationships among the taxa.

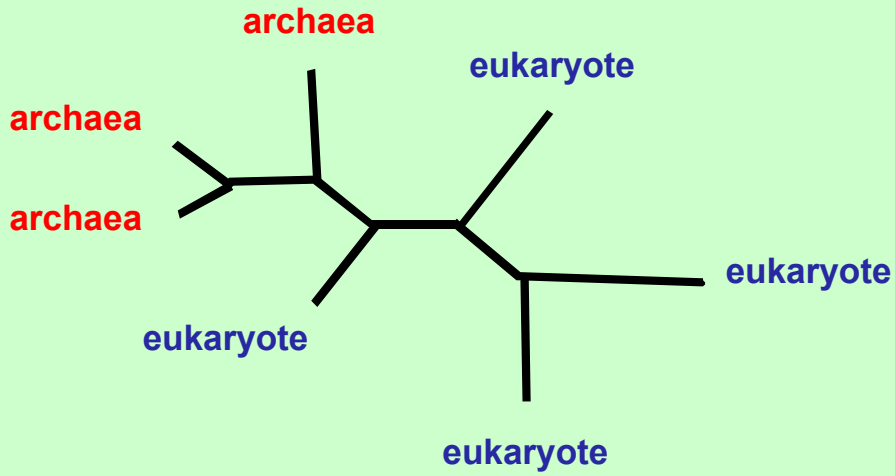
rooted



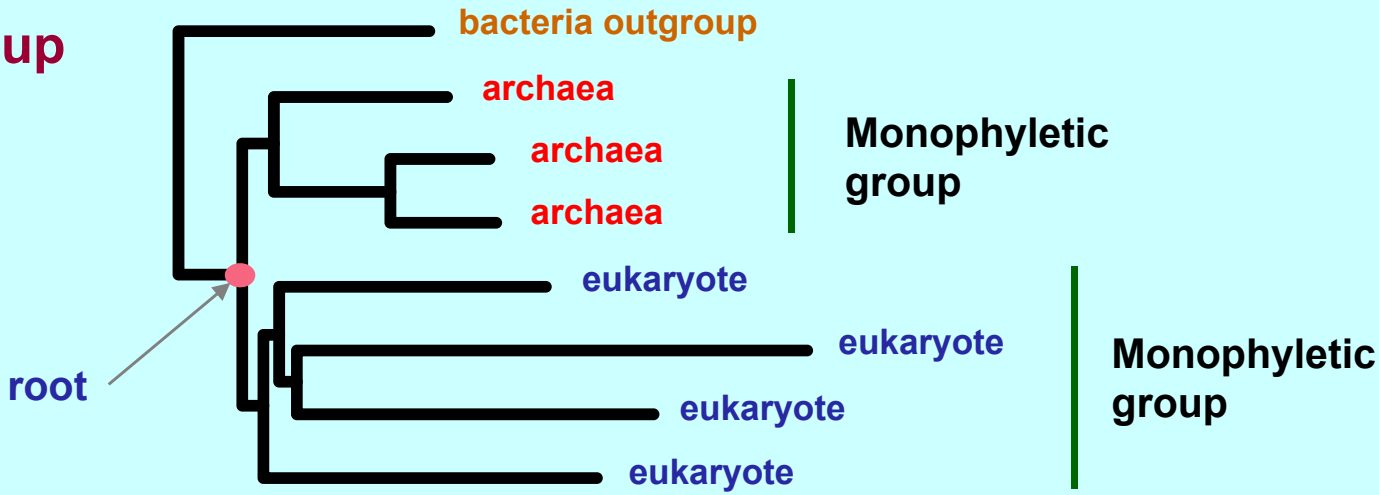
root (R) is common ancestor of all OTUs (operational taxonomic unit)
path from root to OTUs specifies time knowledge of outgroup required to define root

Rooting using an outgroup

unrooted tree

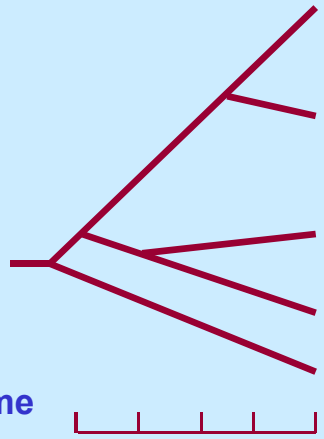


rooted by outgroup

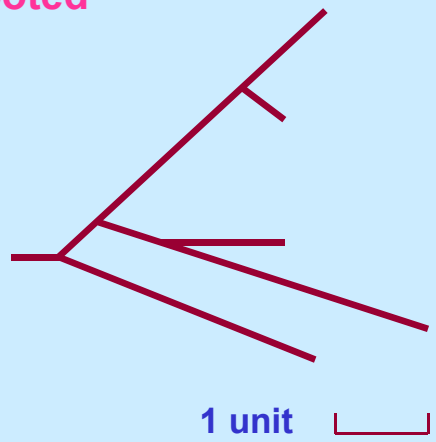


Different visual representations of phylogram trees

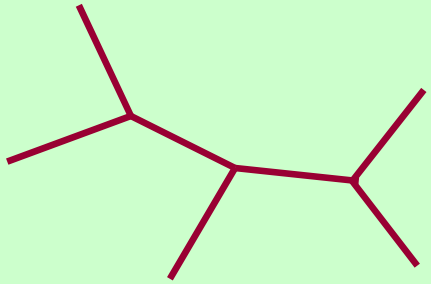
slanted cladogram
rooted



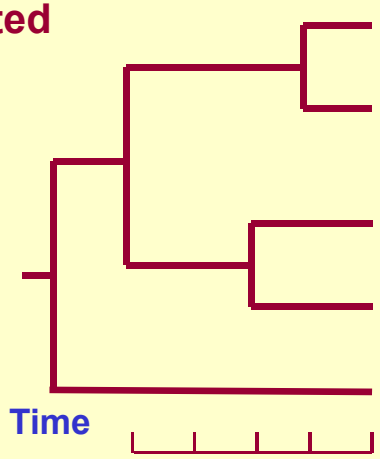
scaled branches
rooted



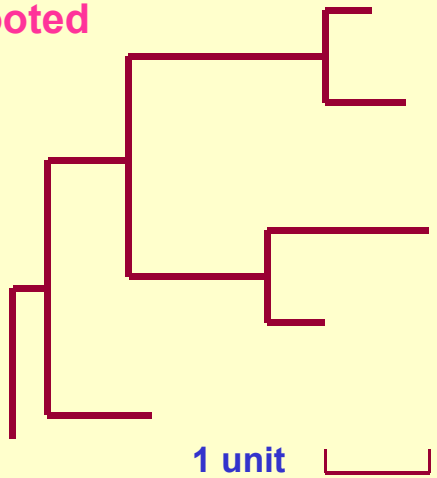
unscaled cladogram
unrooted



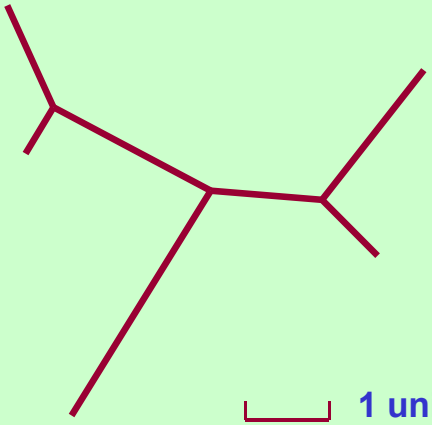
rectangular cladogram
rooted



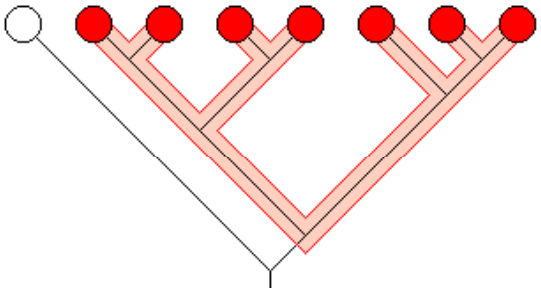
scaled branches
rooted



scaled branches
unrooted

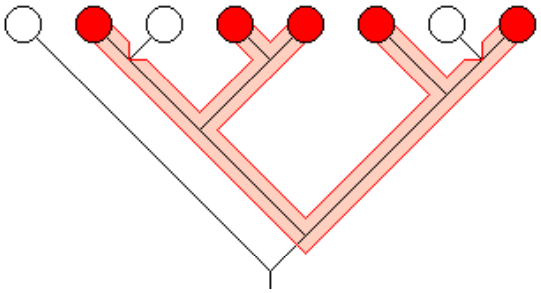


Monophyletic taxon (clade) :



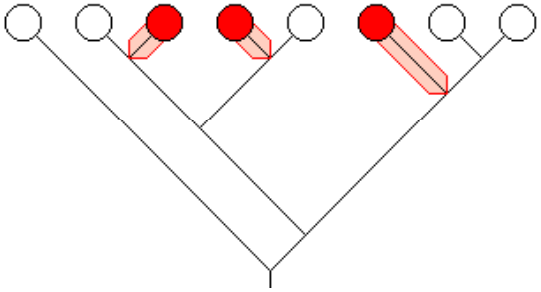
Monophyletic taxon : A group composed of a collection of organisms, including the most recent common ancestor of all those organisms and all the descendants of that most recent common ancestor. *A monophyletic taxon is also called a clade*. **Examples :** Mammalia, Aves (birds), angiosperms, insects, fungi, etc.

Paraphyletic taxon :



Paraphyletic taxon : A group composed of a collection of organisms, including the most recent common ancestor of all those organisms. Unlike a monophyletic group, a paraphyletic taxon does not include all the descendants of the most recent common ancestor. **Examples :** Traditionally defined Dinosauria, fish, gymnosperms, invertebrates, protists, etc.

Polyphyletic taxon :



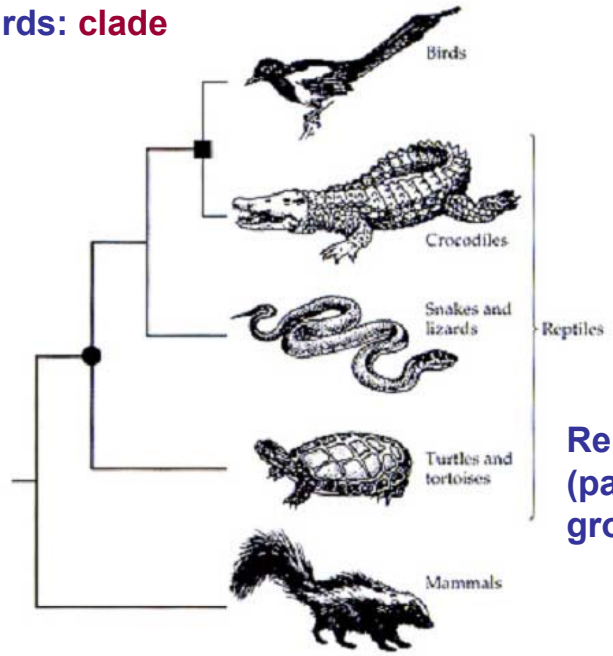
Polyphyletic taxon : A group composed of a collection of organisms in which the most recent common ancestor of all the included organisms is not included, usually because the common ancestor lacks the characteristics of the group. Polyphyletic taxa are considered "unnatural", and usually are reclassified once they are discovered to be polyphyletic. **Examples :** marine mammals, bipedal mammals, flying vertebrates, trees, algae, etc.

Clade vs. Grade

Clade: monophyletic group

Grade: non-monophyletic group, put together out of tradition or convenience, or to reflect morphologically distinct traits

Birds: clade

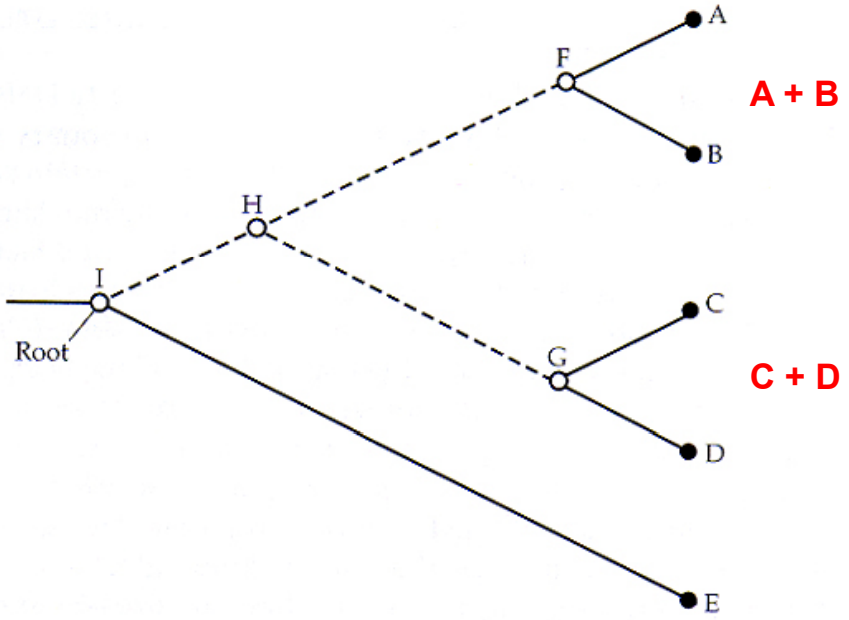


Reptiles: grade (paraphyletic group)

Mammals: clade

Sister Taxa

Sister Taxa: two taxa (= named group of organisms) that are more closely related to each other than either is to a 3rd taxon, and derived from a common ancestral node.



A + B

C + D

Default assumptions in phylogenetics

1. The sequence is correct and originates from the specified source.
2. The sequences are homologous (i.e., are all descended in some way from a shared ancestral sequence).
3. Each position in a sequence alignment is homologous with every other in that alignment.
4. Each of the multiple sequences included in a common analysis has a common phylogenetic history with the others (e.g., there are no mixtures of nuclear and organellar sequences).
5. The sampling of taxa is adequate to resolve the problem of interest.
6. Sequence variation among the samples is representative of the broader group of interest.
7. The sequence variability in the sample contains phylogenetic signal adequate to resolve the problem of interest.

Additional assumptions in phylogenetics

1. The sequences in the sample evolved according to a single stochastic process.
2. All positions in the sequence evolved according to the same stochastic process.
3. Each position in the sequence evolved independently.

Homologs

orthologs/orthologous (直向同源)：

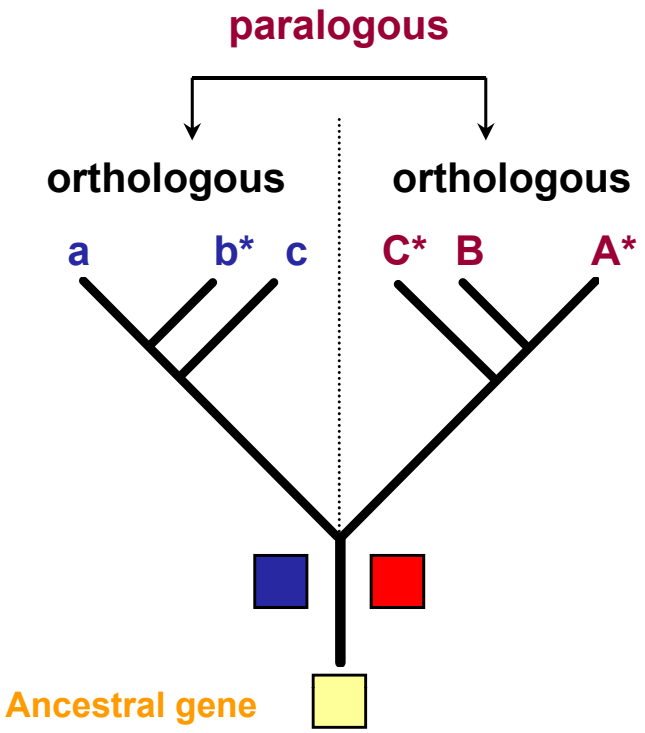
共同祖先的直接後代(沒有發生基因複製事件)之間的同源基因稱為直向同源。

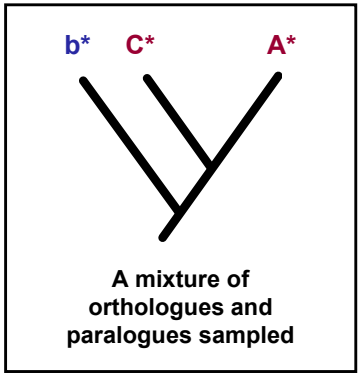
Orthologs are homologs produced by speciation.

paralogs/paralogous (共生同源)：

兩個物種 A 和 B 的同源基因，分別是共同祖先基因組中由複製事件而產生的不同拷貝的後代，這被稱為共生同源基因。

Paralogs are homologs produced by gene duplication.





Duplication to give 2 copies = paralogues on the same genome

Xenologs **are homologs resulting from horizontal gene transfer between two organisms.**

Synologs **are homologs resulting from genes ended up in one organism through fusion of lineages**

PHYLOGENETIC DATA ANALYSIS: THE FOUR STEPS

A straightforward phylogenetic analysis consists of four steps:

Alignment

- Building the data model
- Extraction of a phylogenetic data set

1

Determining the substitution model

- Substitution rates between bases
- Among-site substitution rate heterogeneity
- Substitution rates between amino acids

2

Tree building

3

Distance-Based Methods

1. Unweighted Pair Group Method with Arithmetic Mean (UPGMA).
2. Neighbor Joining (NJ).
3. Fitch-Margoliash (FM).
4. Minimum Evolution (ME).

Character-Based Methods

1. Maximum Parsimony (MP).
2. Maximum Likelihood (ML).

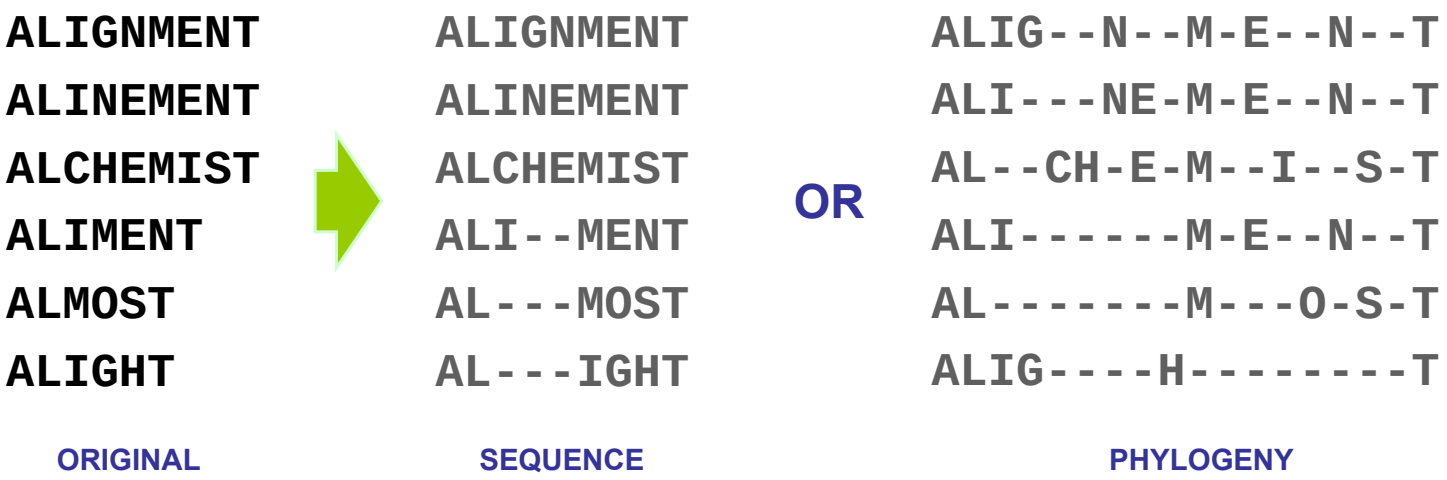
4

Tree evaluation

- Randomized Trees (Skewness Test)
- Randomized Character Data (Permutation Tests)
- Bootstrap
- Likelihood Ratio Tests

1 Alignment

Aligned sequence positions subjected to phylogenetic analysis represent a priori phylogenetic conclusions because the sites themselves (not the actual bases) are effectively assumed to be genealogically related, or homologous. Steps in building the alignment include selection of the alignment procedure(s) and extraction of a phylogenetic data set from the alignment.



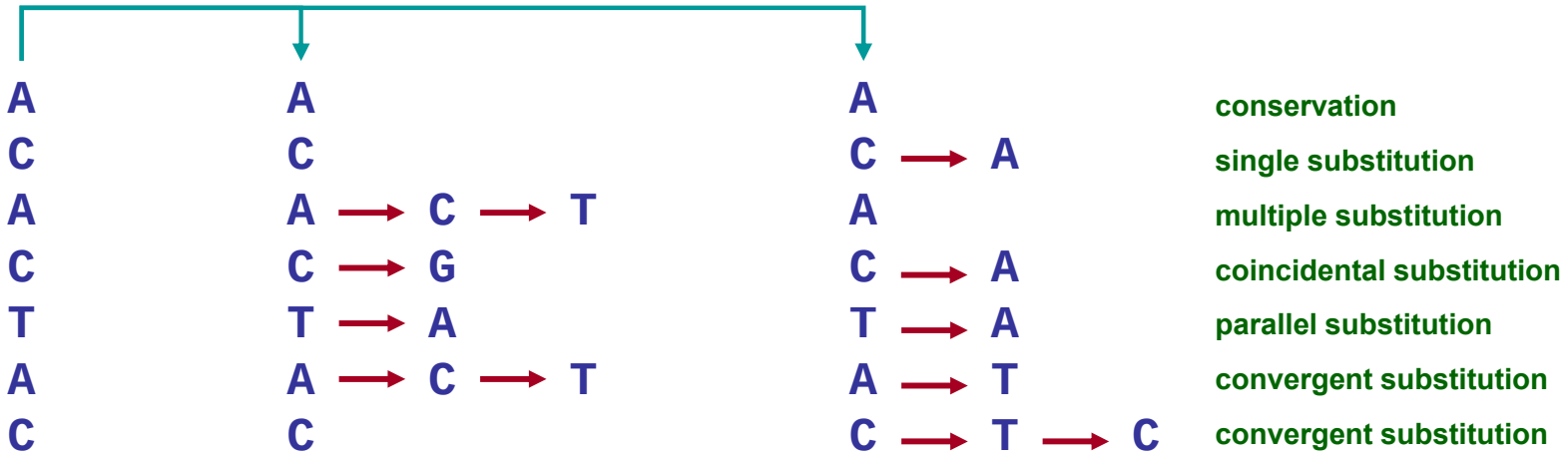
Notices of multiple sequence alignment

- The alignment step in phylogenetic analysis is one of the most important because it produces the data set on which models of evolution are used.
- It is not uncommon to edit the alignment, deleting unambiguously aligned regions and inserting or deleting gaps to more accurately reflect probable evolutionary processes that led to the divergence between sequences.
- It is useful to perform phylogenetic analyses based on a series of slightly modified alignments to determine how ambiguous regions in the alignment affect the results and what aspects of the results one may have more or less confidence in.

❷ Modeling

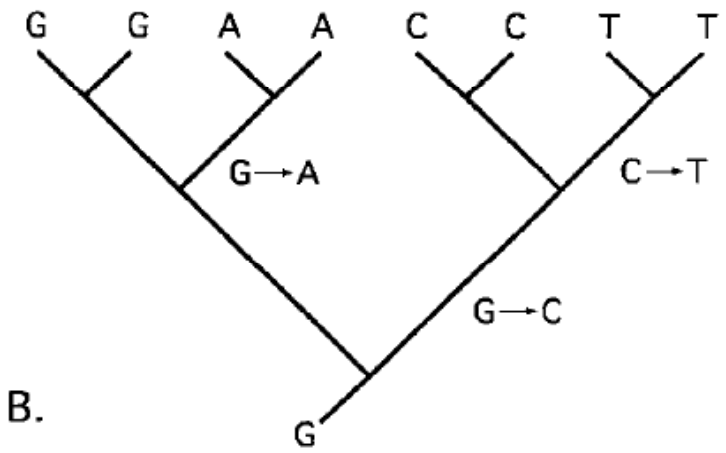
In general, substitutions are more frequent between bases that are biochemically more similar.

In the case of DNA, the four types of transition (A → G, G → A, C → T, T → C) are usually more frequent than the eight types of transversion (A → C, A → T, C → G, G → T, and the reverse). Such biases will affect the estimated divergence between two sequences.

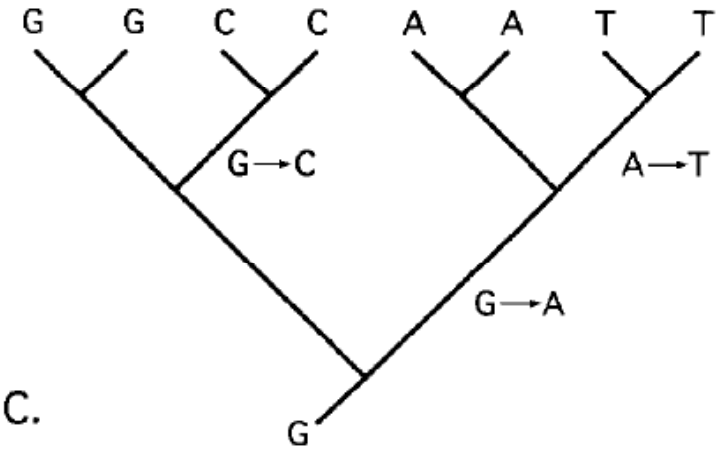


Character-state weight matrices have usually been estimated more or less by eye, but they can also be derived from a rate matrix. For example, if it is presumed that each of the two transitions occurs at double the frequency of each transversion, a weight matrix can simply specify, for example, that the cost of A-G is 1 and the cost of A-T is 2.

	A	C	G	T
A	-	2	1	2
C	2	-	2	1
G	1	2	-	2
T	2	1	2	-



B.



C.

Specification of the relative rates of substitution among particular residues usually takes the form of a square matrix; the number of rows/columns is four in the case of bases, 20 in the case of amino acids (e.g., in PAM and BLOSUM matrices), and 61 in the case of codons (excluding stop codons).

The PAM 250 scoring matrix

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	2																			
R	-2	6																		
N	0	0	2																	
D	0	-1	2	4																
C	-2	-4	4	-5	4															
Q	0	1	1	2	-5	4														
E	0	-1	1	3	-5	2	4													
G	1	-3	0	1	-3	-1	0	5												
H	-1	2	2	1	-3	3	1	-2	6											
I	-1	-2	-2	-2	-2	-2	-2	-3	-2	5										
L	-2	-3	-3	-4	-6	-2	-3	-4	-2	2	6									
K	-1	3	1	0	-5	1	0	-2	0	-2	-3	5								
M	-1	0	-2	-3	-5	-1	-2	-3	-2	2	4	0	6							
F	-4	-4	-4	-6	-4	-5	-5	-5	-2	1	2	-5	0	9						
P	1	0	-1	-1	-3	0	-1	-1	0	-2	-3	-1	-2	-5	6					
S	1	0	1	0	0	-1	0	1	-1	-1	-3	0	-2	-3	1	3				
T	1	-1	0	0	-2	-1	0	0	-1	0	-2	0	-1	-2	0	1	3			
W	-6	2	-4	-7	-8	-5	-7	-7	-3	-5	-2	-3	-4	0	-6	-2	-5	17		
Y	-3	-4	-2	-4	0	-4	-4	-5	0	-1	-1	-4	-2	7	-5	-3	-3	0	10	
V	0	-2	-2	-2	-2	-2	-2	-1	-2	4	2	-2	2	-1	-1	-1	0	-6	-2	4

Distance Matrix Methods

Convert sequence data into a set of discrete pairwise distance values, arranged into a matrix.

- **Distance methods fit a tree to this matrix.**
- **The phylogenetic topology tree is constructed by using a cluster analysis method (like UPGMA or NJ methods).**
- **The phylogeny makes an estimation of the distance for each pair as the sum of branch lengths in the path from one sequence to another through the tree.**

❸ Tree building

Distance - Based Methods

距離建樹方法根據一些尺度計算出雙重序列的距離，然後拋開真實資料，只是根據固定的距離建立進化樹。
這個簡單的運算法，在不同分支的演化速度相近時，可以用來建立親緣樹。因為在上述假設之下，核甘酸或胺基酸的置換速率與親緣遠近大約成正比，所以使用算術平均數來表示距離還算合理。此法採用一系列漸進的雙序列並列分析來做。在程式啟動後，會先將各序列兩兩比對，以找出未來做進一步並列的順序。原則上是先將最相似的序列排列在一起，變為一群(cluster)，然後再將剩餘序列中與這兩個序列最相似的一個，與這兩個排好的序列群做並列分析。最常用的基於特徵符的建樹方法包括 UPGMA 和 NJ。

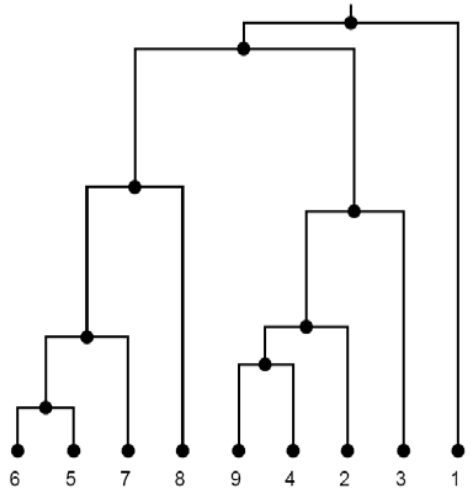
Character - Based Methods

基於特徵符的建樹方法在建立進化樹時，優化了每一個特徵符的真實資料模式的分佈，於是雙重序列的距離不再固定，而是取決於進化樹的拓撲結構。最常用的基於特徵符的建樹方法包括 MP 和 ML。

Unweighted Pair Group Method with Arithmetic Mean (UPGMA)

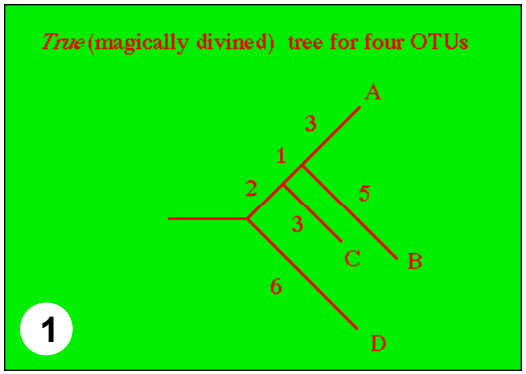
UPGMA 是一種聚類或者說是分類方法；它按照配對序列的最大相似性和連接配對的平均值的標準將進化樹的樹枝連接起來。它還不是一種嚴格的進化距離建樹方法。只有當序列分歧是基於一個分子鐘或者近似等於原始的序列差異性的時候，我們才會期望 UPGMA 會產生一個擁有真實的樹枝長度的準確的拓撲結構。

UPGMA is a clustering or phenetic algorithm - it joins tree branches based on the criterion of greatest similarity among pairs and averages of joined pairs. It is not strictly an evolutionary distance method. UPGMA is expected to generate an accurate topology with true branch lengths only when the divergence is according to a molecular clock or approximately equal to raw sequence dissimilarity. As mentioned earlier, these conditions are rarely met in practice.



UPGMA

UPMGA Tree



2

OTU	A	B	C	D
A	—	8	7	12
B		—	9	14
C			—	11
D				—

Dist. fr A-C to B = $\frac{8 + 9}{2}$ = 8.5 = $\frac{(A \text{ to } B) + (C \text{ to } B)}{2}$

Dist. fr A-C to D = $\frac{12 + 11}{2}$ = 11.5 = $\frac{(A \text{ to } D) + (C \text{ to } D)}{2}$

3

OTU	A-C	B	D
A-C	—	8.5	11.5
B		—	14
D			—

Dist. fr A-C-B to D = $\frac{12 + 14 + 11}{3}$ = 12.33333

= $\frac{(A \text{ to } D) + (B \text{ to } D) + (C \text{ to } D)}{3}$

First node unites A & C with branch lengths of 7/2 = 3.5

Second node unites the A-C clade with B with branch length of 8.5/2 = 4.25

Third node unites A-C-B with D with branch length of 12.33/2 = 6.17

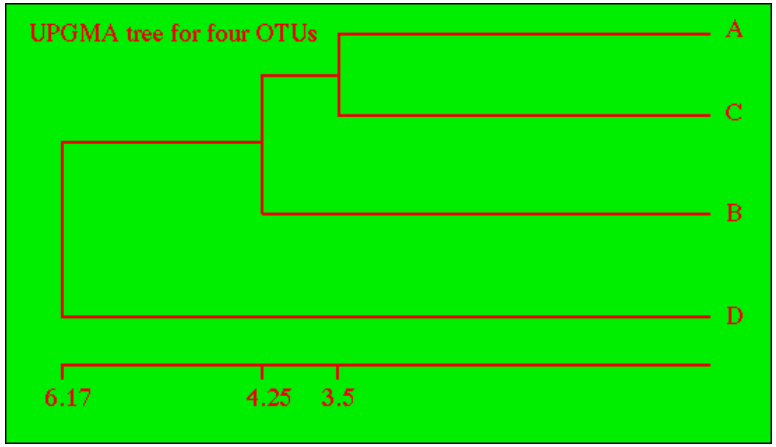
Internode distances can be calculated by subtraction

Node 1 to Node 2 = (Node 2 to B) - ("Height" of Node 1)
= 4.25 - 3.5 = 0.75

"Height" of Node 1 can be taken from EITHER branch length 1-A or 1-C because branch lengths from any node to tip are equal by definition

Node 2 to Node 3 = (Node 2 to D) - ("Height of Node 2)
= 6.17 - 4.25 = 1.91667

4

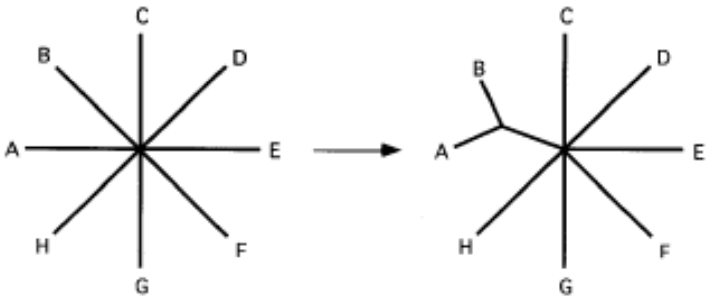


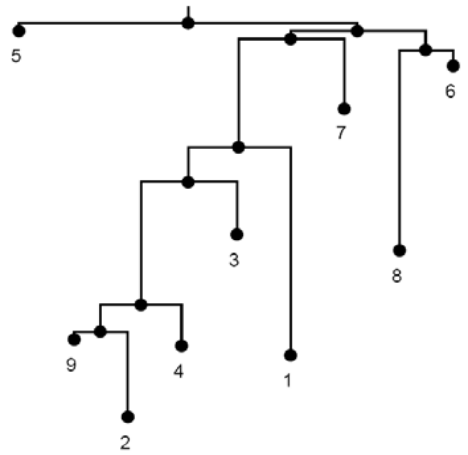
5 http://www.dina.dk/~sestoft/bsa/Match7Applet.html

Neighbor Joining (NJ)

NJ 在距離建樹中經常會用到，不會理會使用什麼樣的優化標準。解析出的進化樹是通過對完全沒有解析出的“星型”進化樹進行“分解”得到，分解的步驟是連續不斷地在最接近（實際上，是最孤立的）的序列對中插入樹枝，而保留進化樹的終端。最接近的序列對被鞏固了，而“星型”進化樹被改善了，這個過程將不斷重複。

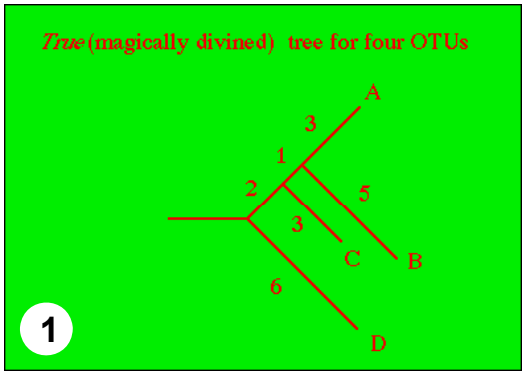
The neighbor-joining algorithm is commonly applied with distance tree building, regardless of the optimization criterion. The fully resolved tree is “decomposed” from a fully unresolved “star” tree by successively inserting branches between a pair of closest (actually, most isolated) neighbors and the remaining terminals in the tree. The closest neighbor pair is then consolidated, effectively reforming a star tree, and the process is repeated. The method is comparatively rapid.





NJ

NJ Tree



2

OTU	A	B	C	D	*r*	*r/2*
A	-	8	7	12	27	13.5
B		-	9	14	31	15.5
C			-	11	27	13.5
D				-	37	18.5

8+7+12

8+9+14

Note that we have two new columns to the right. The first column (r) is the sum of the distances from the row OTU to all other OTUs. Thus 8+7+12 = 27 (A to everything else); 8+9+14 = 31 (B to everything else); etc. The r/2 is something we will use later. The denominator (the 2) is the matrix size (number of OTUs) minus two. I will explain that later.

3

OTU	A	B	C	D
A	-	8	7	12
B	-21.00	-	9	14
C	-20.00	-20.00		11
D	-20.00	-20.00	-21.00	-

Original A-B value (8) minus the average of the A and B r-values [(27+31)/2 = 29].

8 - 29 = -21.

A-C = -20. Original A-C value (7) minus average of A and C r-values

[(27+27)/2 = 27]. 7 - 27 = -20.

4

B to Node 1: Original B-A distance divided by two (original distance between the components/2) plus (B's r/2 minus A's r/2) divided by two.

8/2 + (15.5 - 13.5)/2 = 5

B to Node 1 = 5

A to B = 8; B to Node 1 = 5. Therefore A to Node 1 = 8 - 5 = 3.

A to Node 1 = 3

Alternative method starting with A to Node 1:

(Original A to B) + (A's r/2 minus B's r/2) divided by two

8/2 + (13.5 - 15.5)/2 = 4 + -1 = 3

Finally B to Node 1 = A to B - A to Node 1 = 8 - 3 = 5

NJ Tree (cont' 1)

	C	D	Node 1	*r*	*r/1*
C	-	11	4	15	15
D	-6.5	-	9	20	20
Node 1	-10	-7.5	-	13	13

5

C to Node 1. Original C to A (=7) minus A to Node 1 (=3) plus Original C to B (=9) minus B to Node 1 (=5) all divided by two.

So… C to Node 1 = [(7-3) + (9-5)]/2 = 4.

D to Node 1. Original D to A (=12) minus A to Node 1 (=3) plus Original D to B (=14) minus B to Node 1 (=5) all divided by two.

So… D to Node 1 = [(12-3) + (14-5)]/2 = 9.

D to C = Original D to C minus the sum of the (reduced matrix) r-values divided by two.

11-(15+20)/2 = -6.5

Node 1 to C = Original Node 1 to C [N.B., this value comes from the upper-diagonal]

minus the sum of their (reduced matrix) r-values divided by two.

4 -(15+13)/2 = -10

Node 1 to D = Original Node 1 to D minus the sum of their (reduced matrix) r-values divided by two.

9 -(20+13)/2 = -7.5

6

C to Node 2 = (Original C to Node 1)/2 plus (C's r/1 minus Node 1's r/1)/2.

4/2 + (15-13)/2 = 3

C to Node 2 = 3

Node 1 to Node 2 = (Original C to Node 1) minus distance just computed for C to Node 2.

4 - 3 = 1

Node 1 to Node 2 = 1

Alternative starting with Node 1 to Node 2. What do we know about Node 1 to Node 2? We know something that INCLUDES it, which is C to Node 1 (= C to Node 2, which we don't want, plus Node 2 to Node 1, which we do want).

Node 1 to Node 2 = (C to Node 1)/2 plus (Node 1's r/1 - C's r/1)

NJ Tree (cont' 2)

	D	Node 2	r
D	-	8	
Node 2		-	

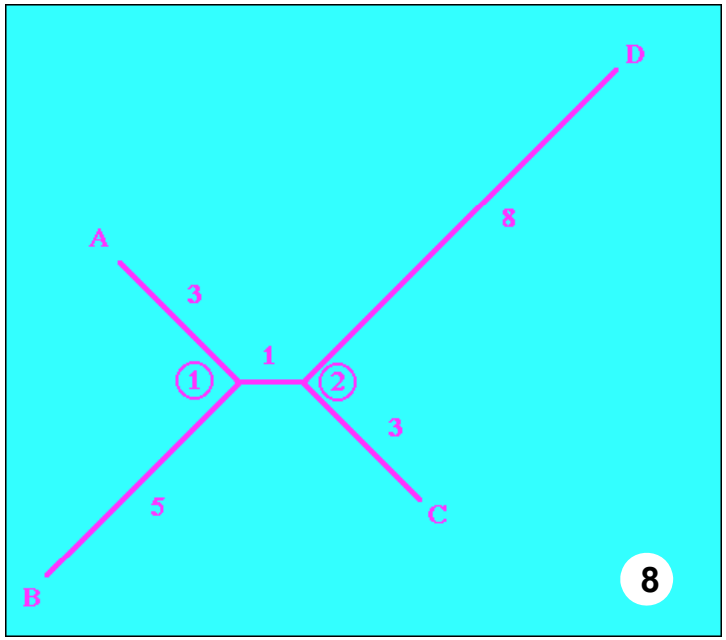
7

D to Node 2 =

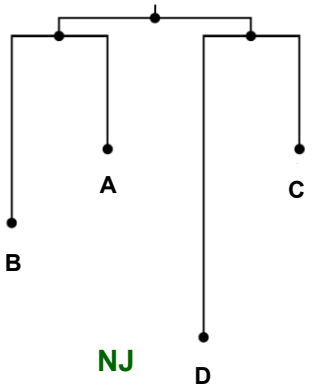
[(D to Node 1 minus Node 1 to Node 2) + (D to C minus C to Node 2)]/2

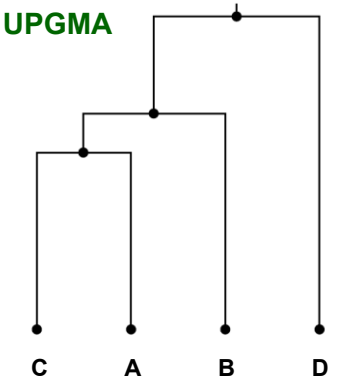
[(9 - 1) + (11-3)]/2 = 8

D to Node 2 = 8



9 http://www.dina.dk/~sestoft/bsa/Match7Applet.html





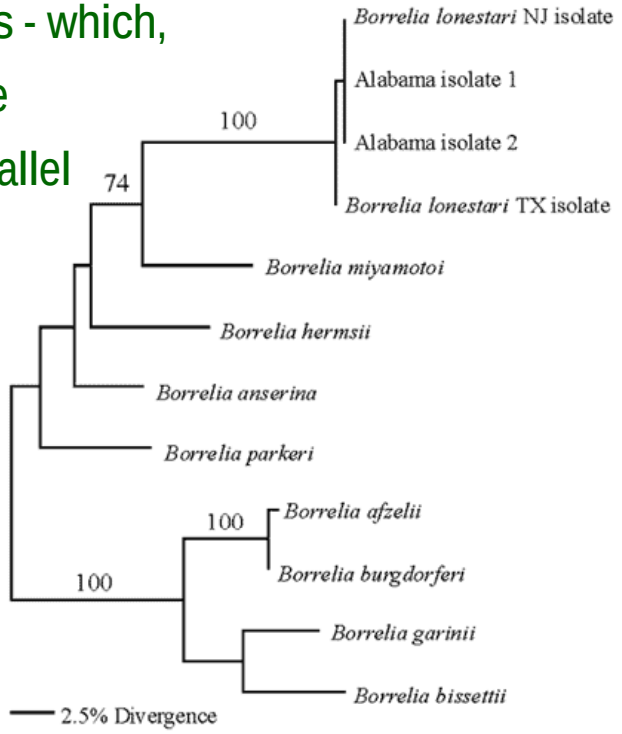
Character Matrix Methods

1. Parsimony is the most popular method for reconstructing ancestral relationships.
2. Parsimony allows the use of all known evolutionary information in tree.
3. The phylogenetic topology tree is constructed by using a cluster analysis method (like MP or ML methods).
4. Approaches involve two components:
    - A search through space of trees.
    - A procedure to find the minimum number of changes needed to explain the data – used for scoring each tree.

Maximum Parsimony (MP)

最大節約方法是一種優化標準，對資料最好的解釋也是最簡單的，而最簡單的所需要的特別假定也最少。在實際應用中，MP 進化樹是最短的；也是變化最少的進化樹，根據定義，這個進化樹的平行變化最少，或者說是同形性最低。MP 中有一些變數與特徵符狀態改變的可行方向不盡相符。

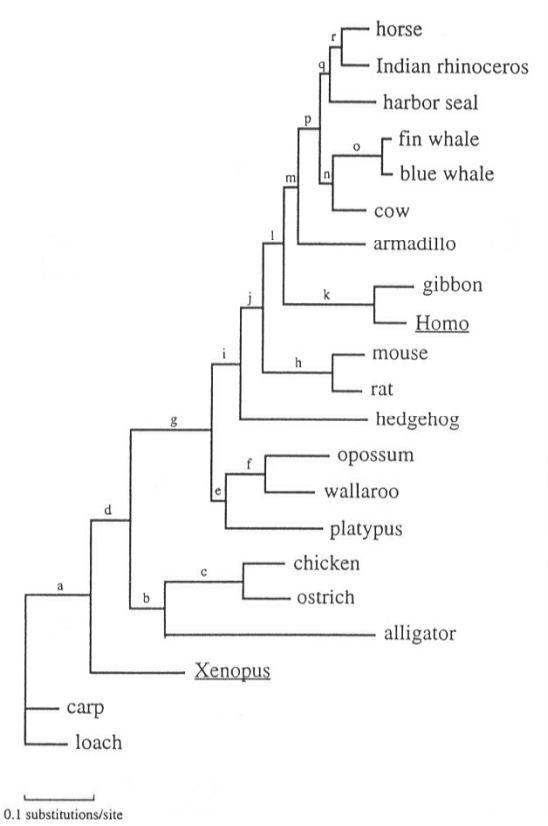
Maximum Parsimony (MP). Maximum parsimony is an optimization criterion that adheres to the principle that the best explanation of the data is the simplest, which in turn is the one requiring the fewest ad hoc assumptions. In practical terms, the MP tree is the shortest - the one with the fewest changes - which, by definition, is also the one with the fewest parallel changes. There are several variants of MP that differ with regard to the permitted directionality of character state change.



Maximum Likelihood (ML)

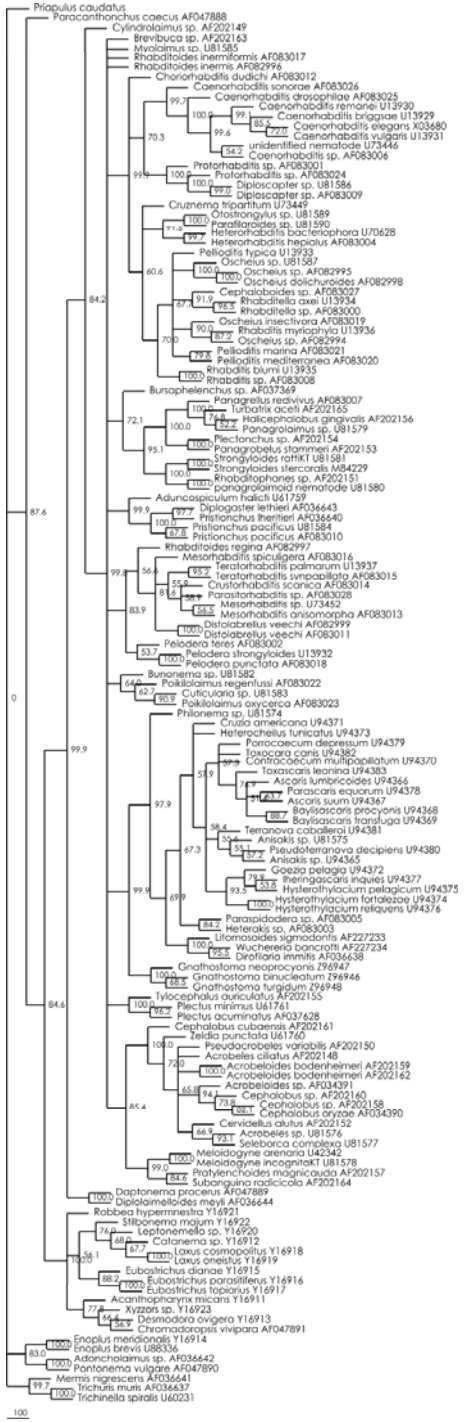
ML 對系統發育問題進行了徹底搜查。ML 期望能夠搜尋出一種進化模型（包括對進化樹本身進行搜索），使得這個模型所能產生的資料與觀察到的資料最相似。

Maximum Likelihood (ML). ML turns the phylogenetic problem inside out. ML searches for the evolutionary model, including the tree itself, that has the highest likelihood of producing the observed data.

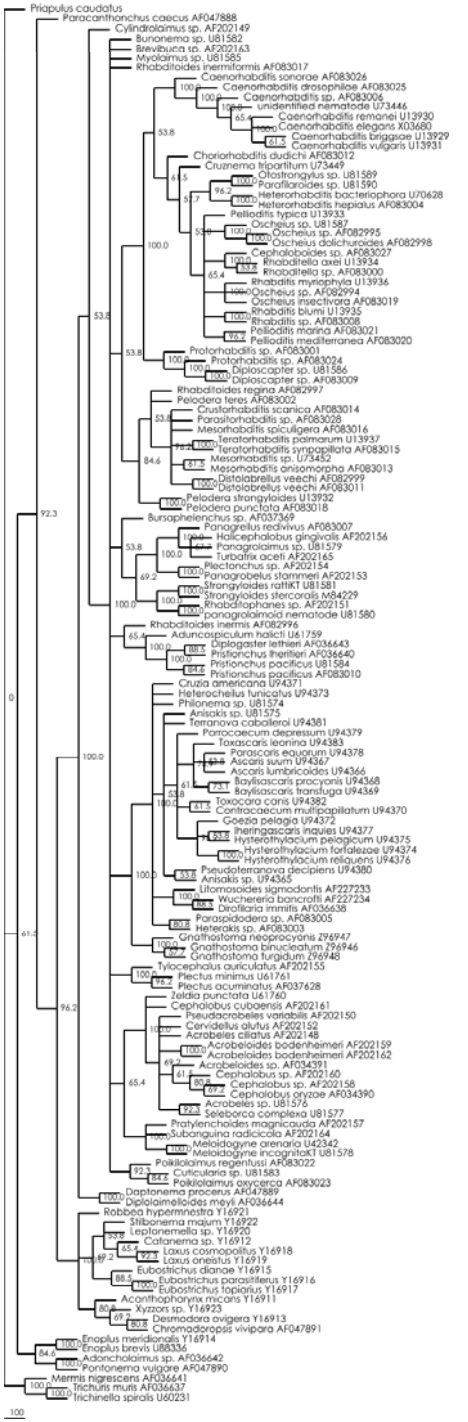


142 nematode
SSU sequences

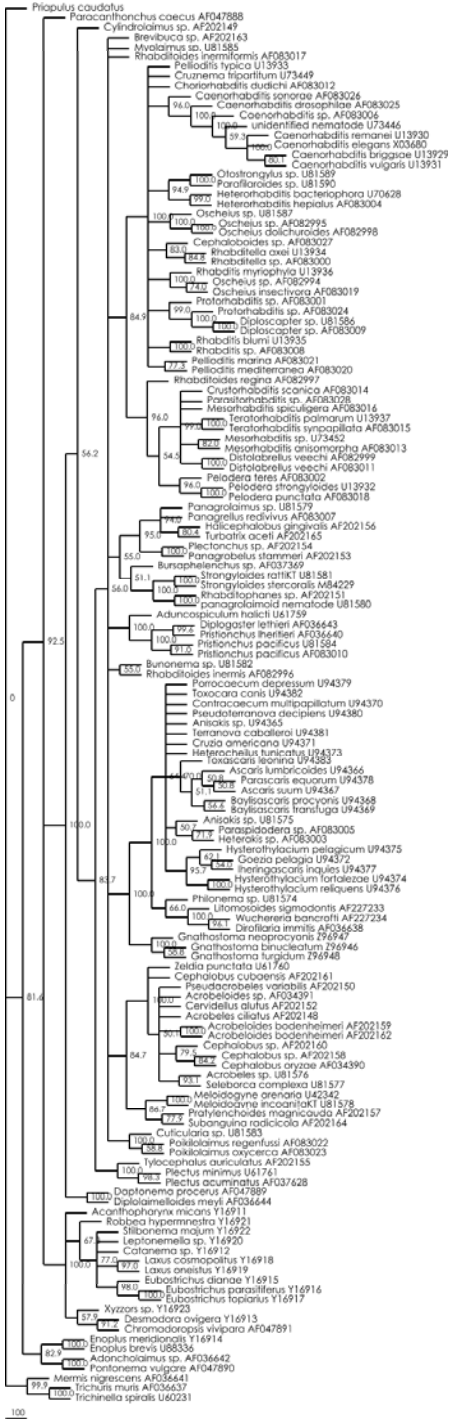
Bootstrap distance tree



Bootstrap maximum likelihood tree



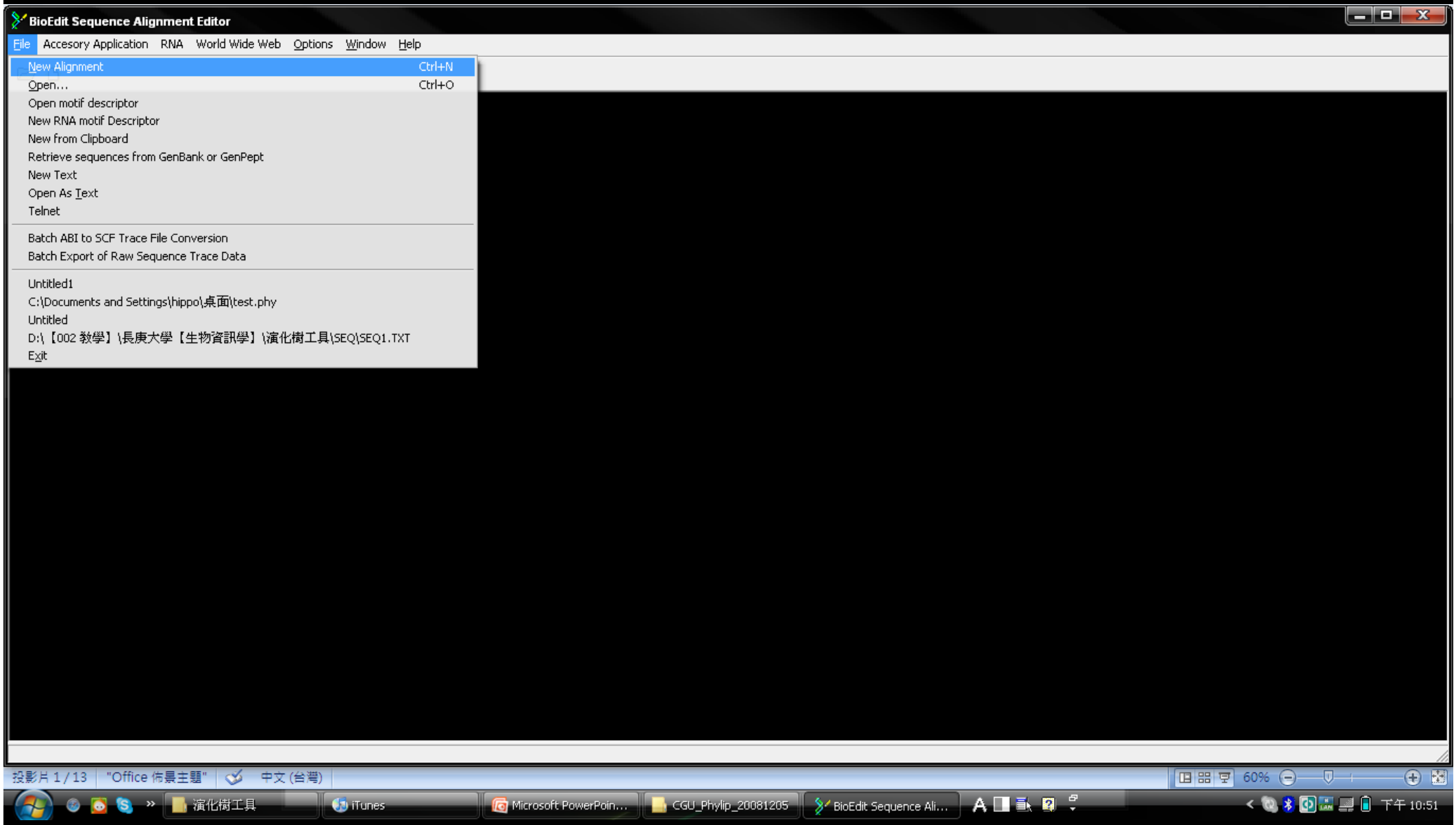
Bootstrap maximum parsimony tree

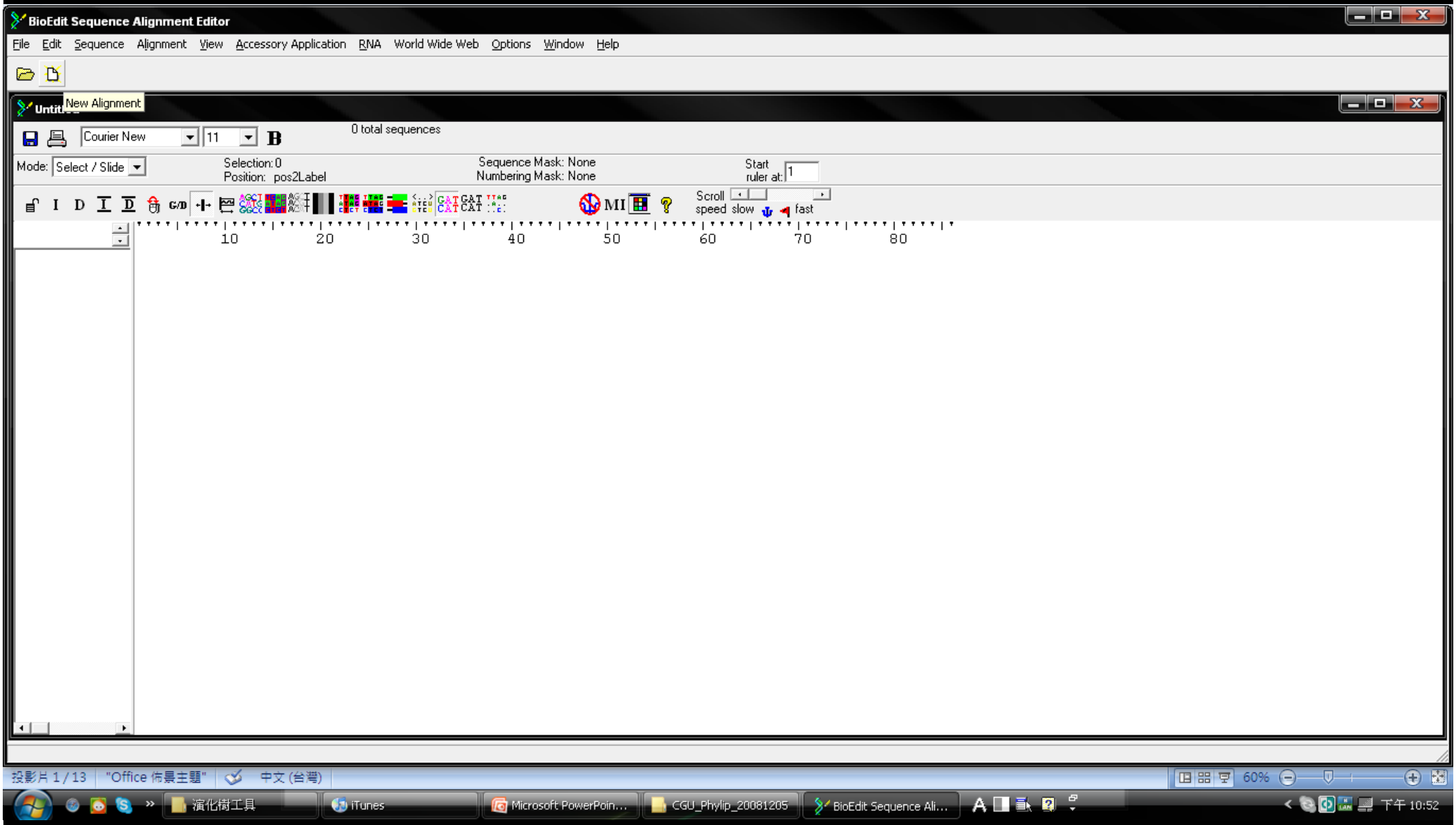


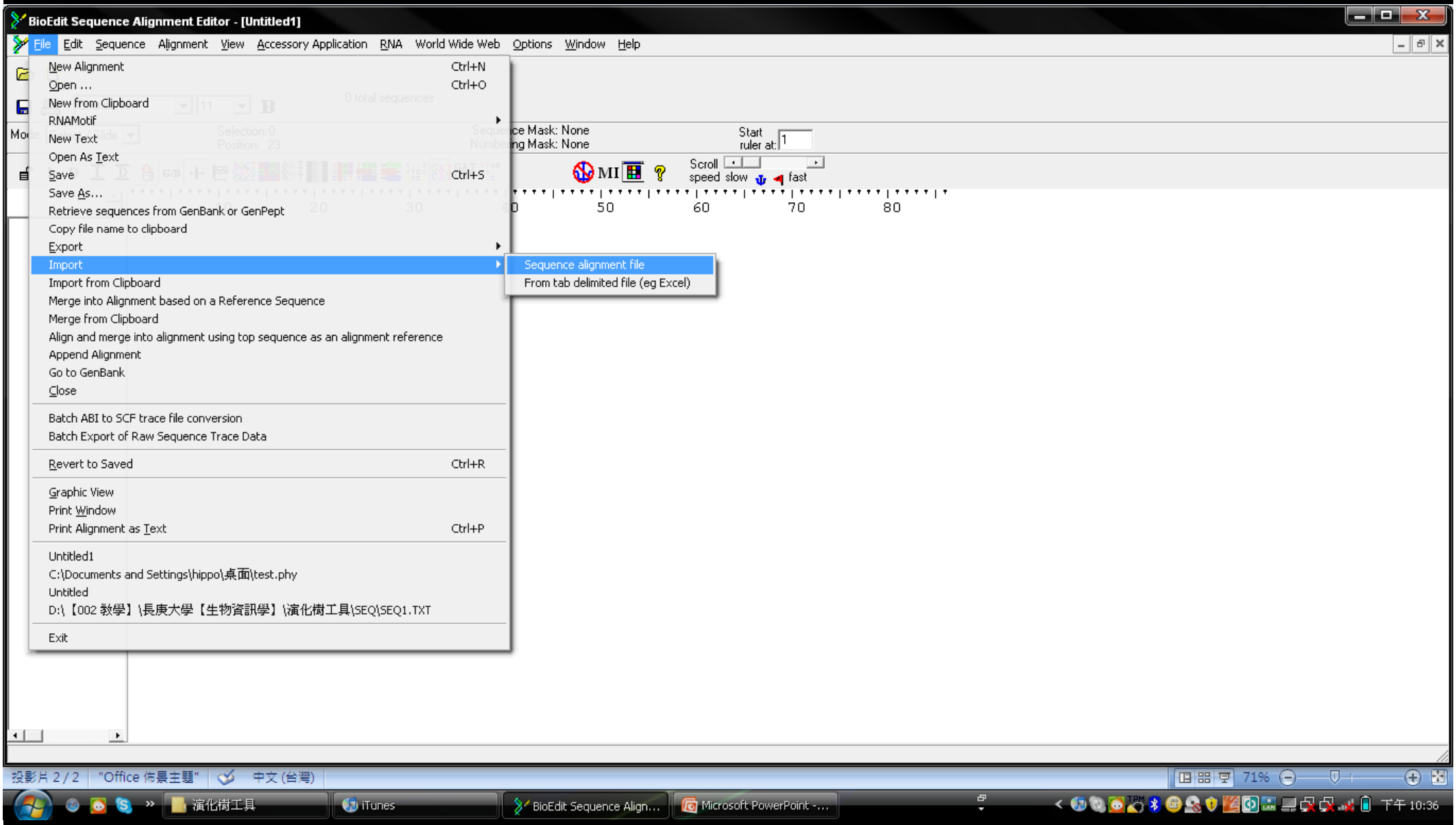
TOOLS

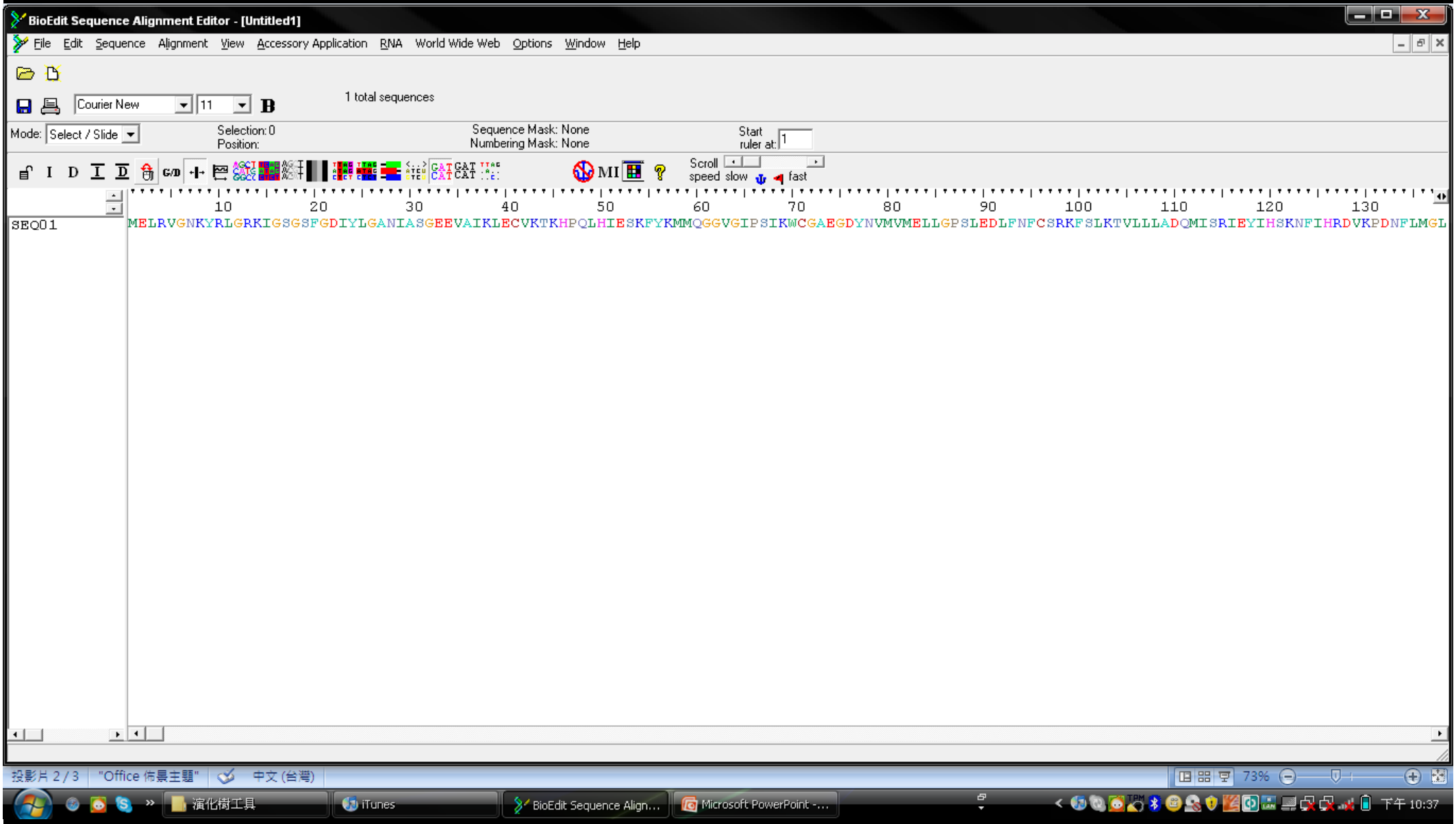
BioEdit	http://www.mbio.ncsu.edu/BioEdit/BioEdit.html
PHYLIP	http://evolution.genetics.washington.edu/phylip/getme.html
TreeView X	http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/
HyperTree	http://kinase.com/tools/HyperTree.html
Java	http://www.java.com/zh_TW/
CLC Sequence Viewer	http://www.clcbio.com/index.php?id=1113
Geneious	http://www.geneious.com/default,83,home.sm

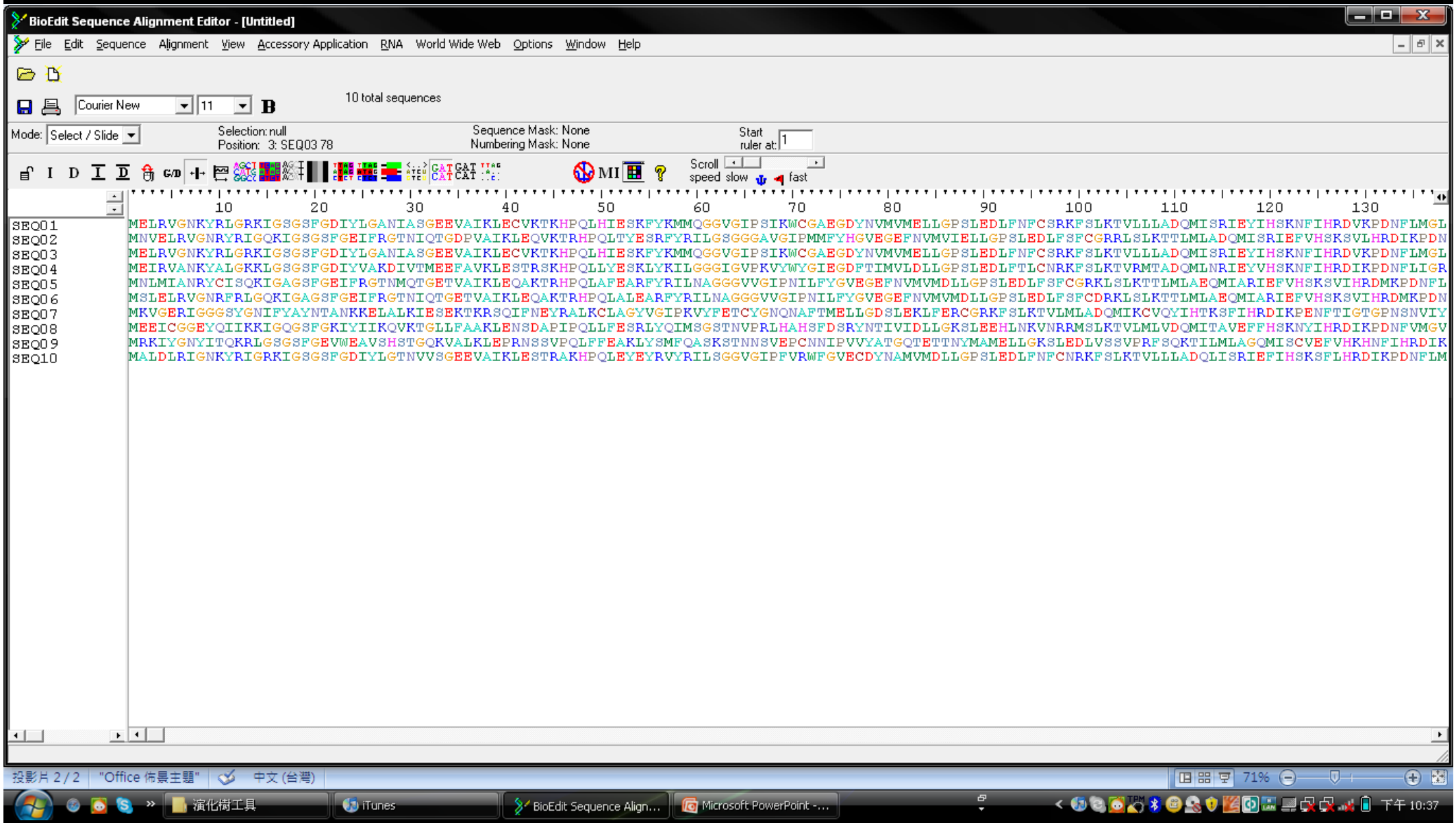
BIOEDIT → ClustalW Multiple Alignment → PHYLIP

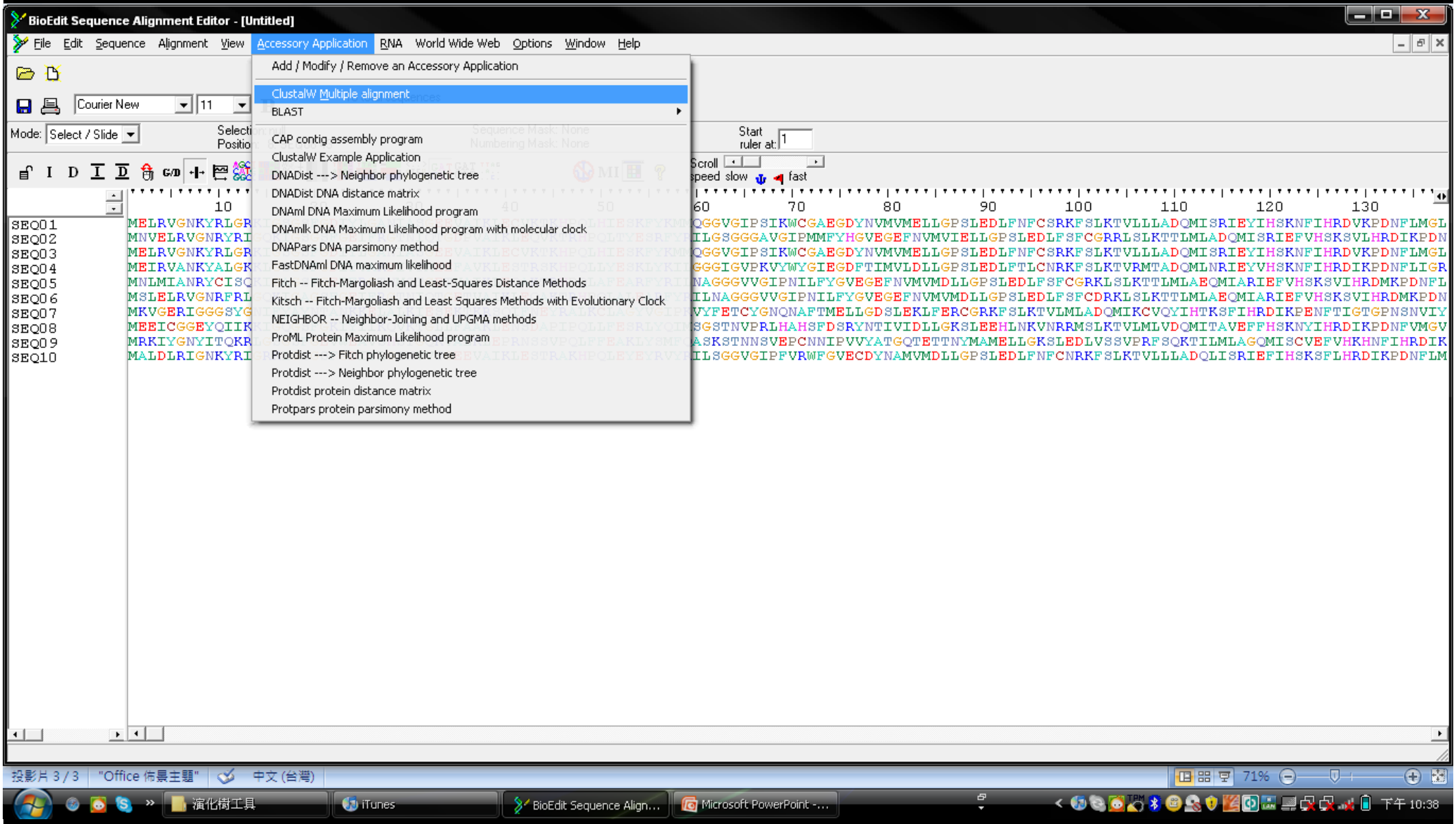


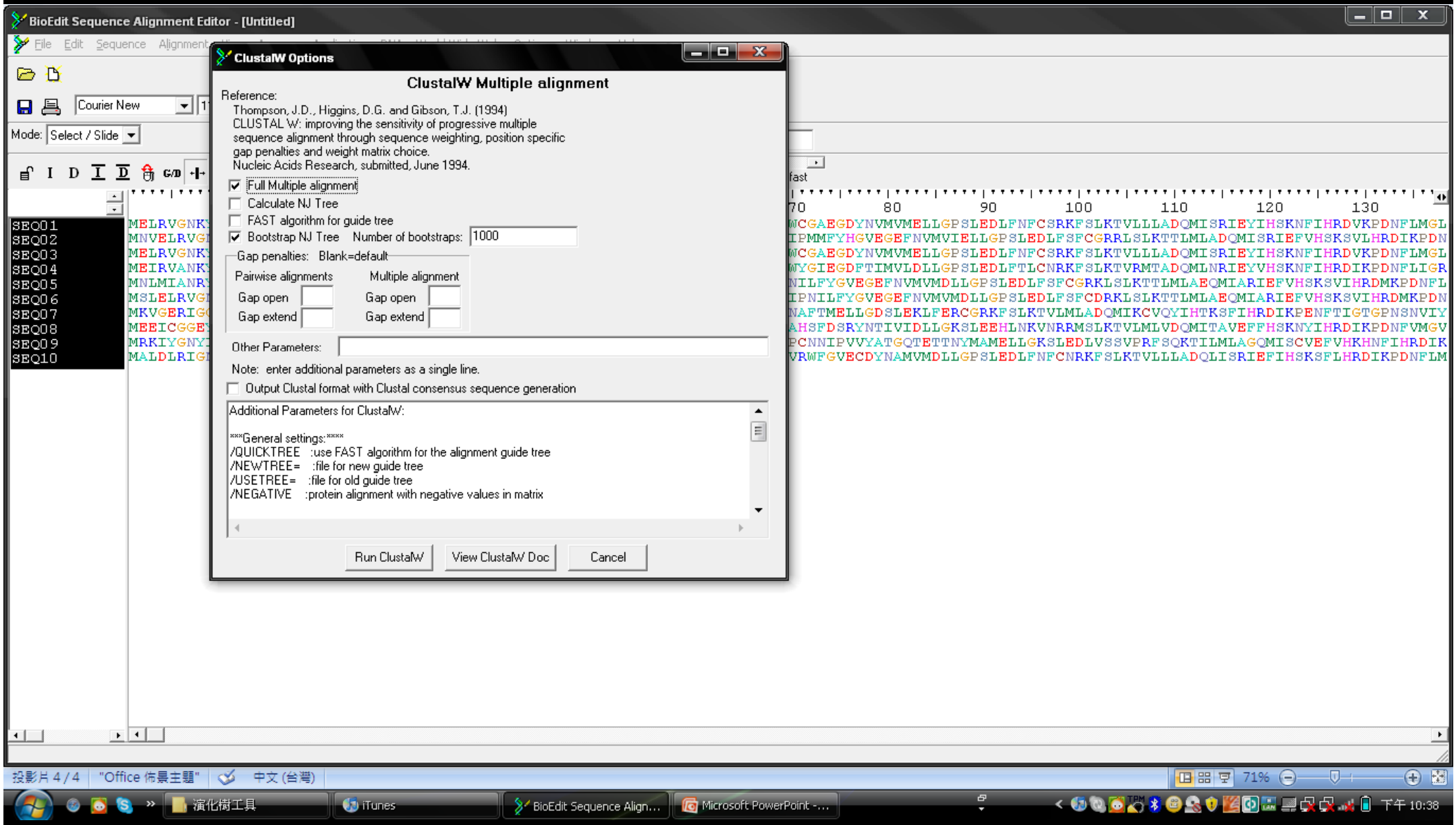

BioEdit Sequence Alignment Editor
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
New Alignment
Courier New
11
0 total sequences
Mode: Select / Slide
Selection: 0
Position: pos2Label
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
10 20 30 40 50 60 70 80
投影片 1 / 13
"Office 佈景主題"
中文 (台灣)
60%
演化樹工具
iTunes
Microsoft PowerPoin...
CGU_Phylip_20081205
BioEdit Sequence Ali...
下午 10:52

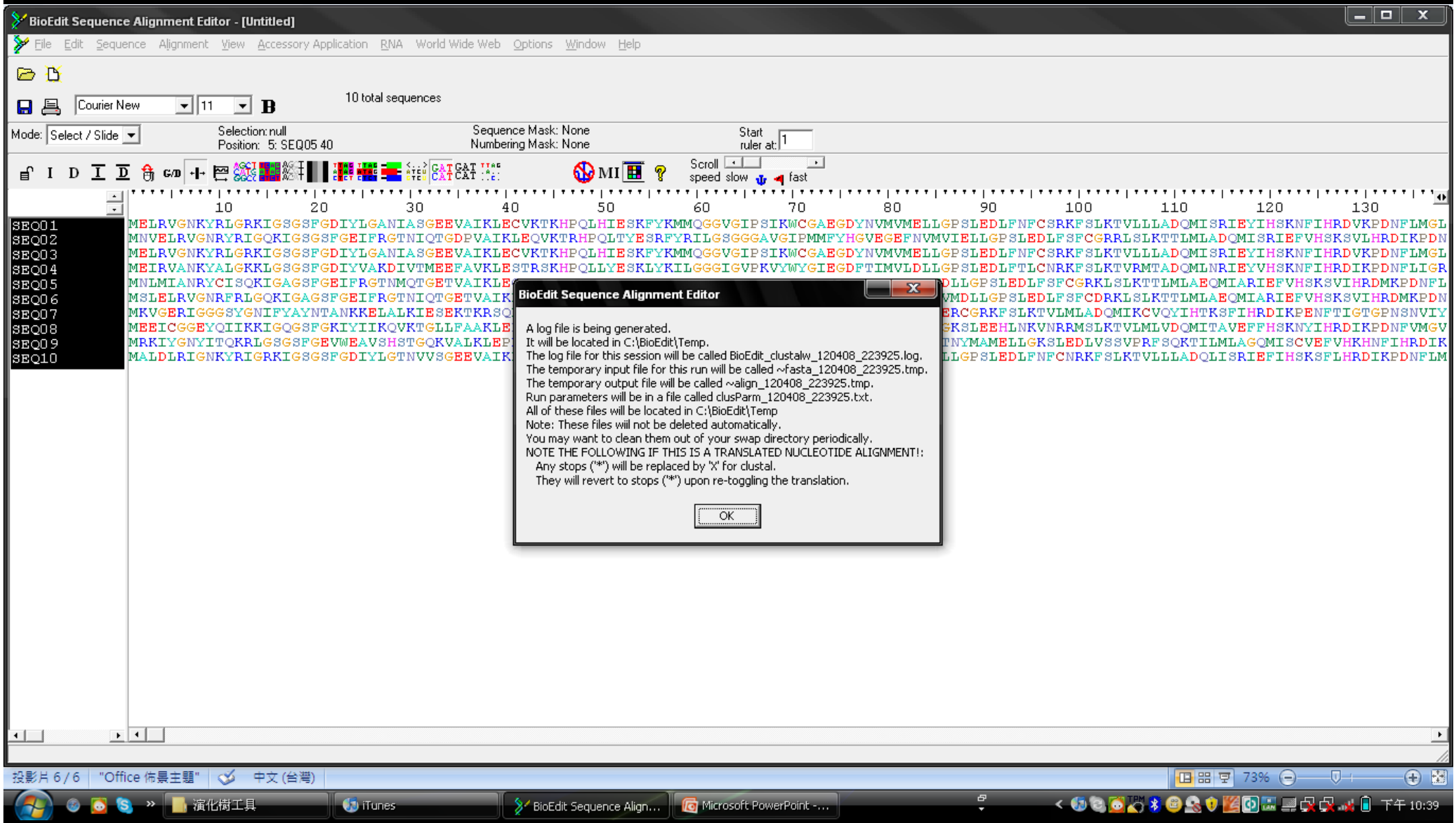

BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
New Alignment Ctrl+N
Open ... Ctrl+O
New from Clipboard
RNAMotif
New Text
Open As Text
Save Ctrl+S
Save As...
Retrieve sequences from GenBank or GenPept
Copy file name to clipboard
Export
Import
Import from Clipboard
Merge into Alignment based on a Reference Sequence
Merge from Clipboard
Align and merge into alignment using top sequence as an alignment reference
Append Alignment
Go to GenBank
Close
Batch ABI to SCF trace file conversion
Batch Export of Raw Sequence Trace Data
Revert to Saved Ctrl+R
Graphic View
Print Window
Print Alignment as Text Ctrl+P
Untitled1
C:\Documents and Settings\hippo\桌面\test.phy
Untitled
D:\【002 教學】\長庚大學【生物資訊學】\演化樹工具\SEQ\SEQ1.TXT
Exit
Sequence alignment file
From tab delimited file (eg Excel)
Start ruler at: 1
Scroll speed slow fast
50
60
70
80
投影片 2 / 2
"Office 佈景主題"
中文 (台灣)
71%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:36

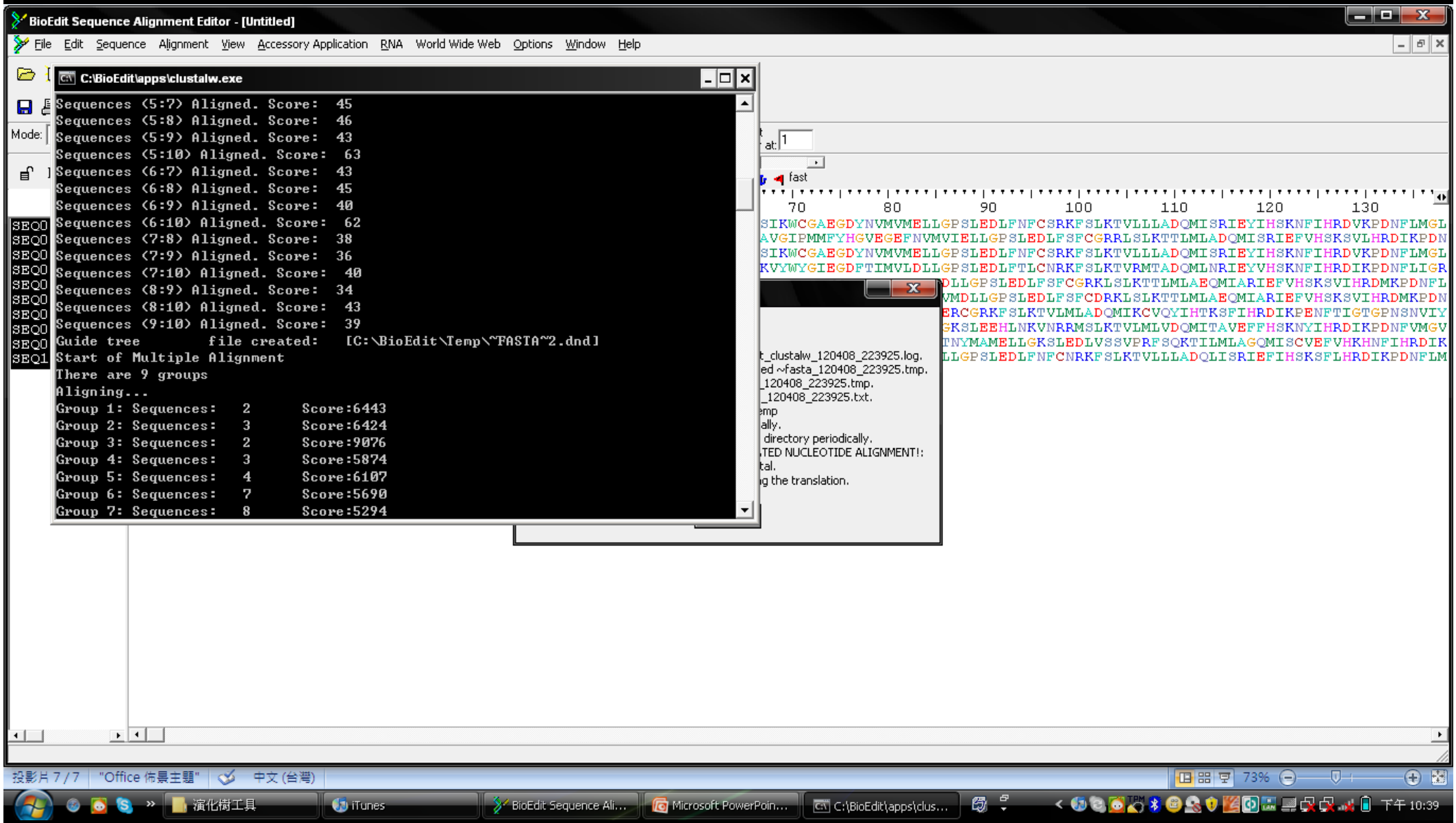

BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
1 total sequences
Mode: Select / Slide
Selection: 0
Position:
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
10 20 30 40 50 60 70 80 90 100 110 120 130
SEQ01
MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGGVGIPSIKWCGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLLADQMISRIEYIHSKNFIHRDVKPDNFLMGL
投影片 2 / 3
"Office 佈景主題"
中文 (台湾)
73%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:37

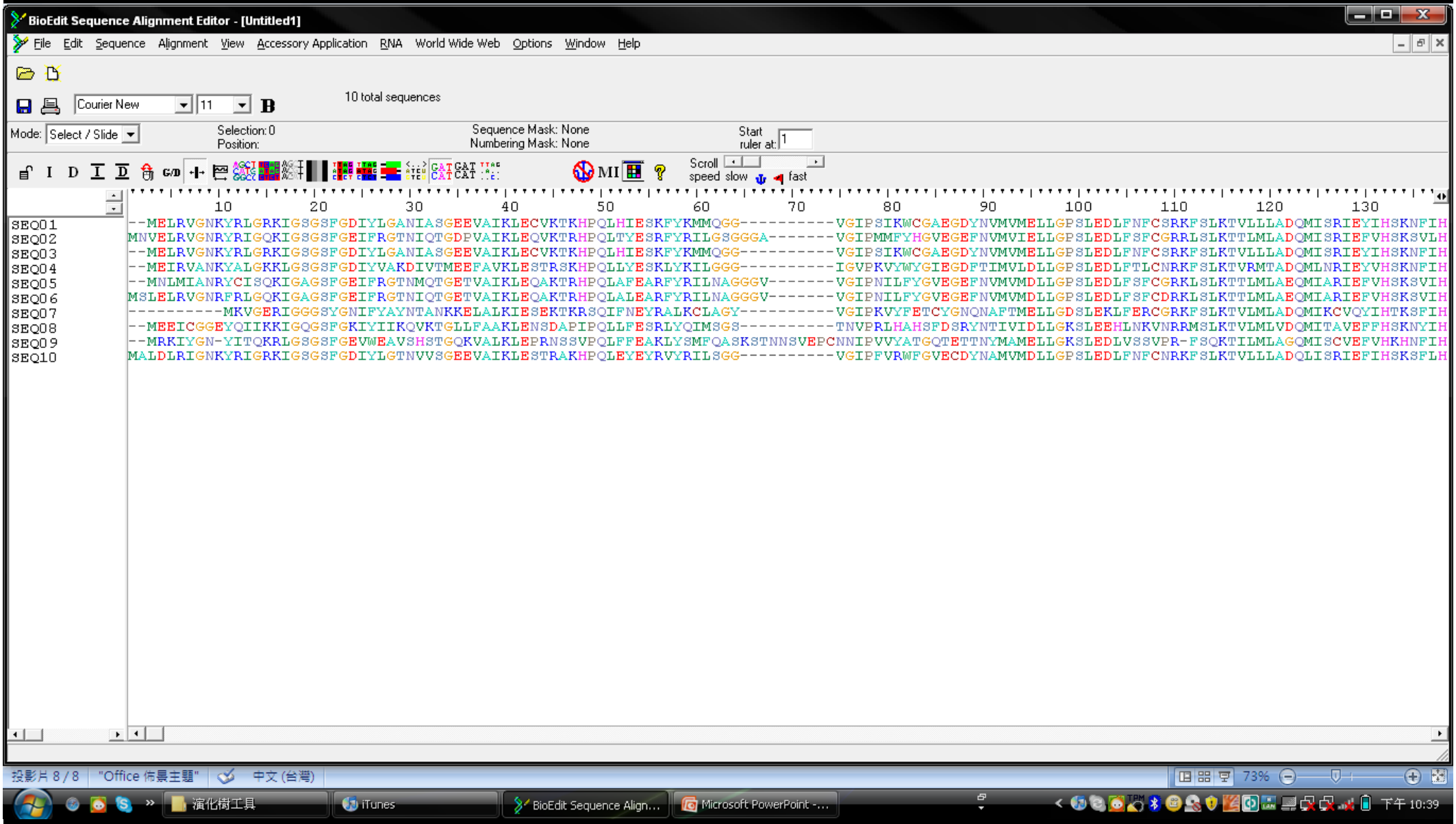

BioEdit Sequence Alignment Editor - [Untitled]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
10 total sequences
Mode: Select / Slide
Selection: null
Position: 3: SEQ03 78
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
10 20 30 40 50 60 70 80 90 100 110 120 130
SEQ01 MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGGVGIPSIKWCGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLLADQMISRIEYIHSKNFIHRDVKPDNFLMGL
SEQ02 MNVELRVGNRYRIGQKIGSGSFGEIFRGTNIQTGDPVAIKLEQVKTRHPQLTYESRFYRILGSGGGAVGIPMMFYHGVEGEFNVMVIELLGPSLEDLFSFCGRRLSLKTTLMLADQMISRIEFVHSKSVLHRDIKPDN
SEQ03 MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGGVGIPSIKWCGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLLADQMISRIEYIHSKNFIHRDVKPDNFLMGL
SEQ04 MEIRVANKYALGKKLGSGSFGDIYVAKDIVTMEEFAVKLESTRSKHPQLLYESKLYKILGGGIGVPKVYWYGIEGDFTIMVLDLLGPSLEDLFTLCNRKFSLKTVRMTADQMLNRIEYVHSKNFIHRDIKPDNFLIGR
SEQ05 MNLMIANRYCISQKIGAGSFGEIFRGTNMQTGETVAIKLEQAKTRHPQLAFEARFYRILNAGGGVVGIPNILFYGVEGEFNVMVMDLLGPSLEDLFSFCGRKLSLKTTLMLAEQMIARIEFVHSKSVIHRDMKPDNFL
SEQ06 MSLELRVGNRFRLGQKIGAGSFGEIFRGTNIQTGETVAIKLEQAKTRHPQLALEARFYRILNAGGGVVGIPNILFYGVEGEFNVMVMDLLGPSLEDLFSFCDRKLSLKTTLMLAEQMIARIEFVHSKSVIHRDMKPDN
SEQ07 MKVGERIGGGSYGNIFYAYNTANKKELALKIESEKTKRSQIFNEYRALKCLAGYVGIPKVYFETCYGNQNAFTMELLGDSLEKLFERCGRKFSLKTVLMLADQMIKCVQYIHTKSFIHRDIKPENFTIGTGPNSNVIY
SEQ08 MEEICGGEYQIIKKIGQGSFGKIYIIKQVKTGLLFAAKLENSDAPIPQLLFESRLYQIMSGSTNVPRLHAHSFDSRYNTIVIDLLGKSLEEHLNKVNRRMSLKTVLMLVDQMITAVEFFHSKNYIHRDIKPDNFVMGV
SEQ09 MRKIYGNYITQKRLGSGSFGEVWEAVSHSTGQKVALKLEPRNSSVPQLFFEAKLYSMFQASKSTNNSVEPCNNIPVVYATGQTETTNYMAMELLGKSLEDLVSSVPRFSQKTILMLAGQMISCVEFVHKHNFIHRDIK
SEQ10 MALDLRIGNKYRIGRKIGSGSFGDIYLGTNVVSGEEVAIKLESTRAKHPQLEYEYRVYRILSGGVGIPFVRWFGVECDYNAMVMDLLGPSLEDLFNFCNRKFSLKTVLLLADQLISRIEFIHSKSFLHRDIKPDNFLM
投影片 2 / 2 "Office 佈景主題" 中文 (台灣)
71%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:37


BioEdit Sequence Alignment Editor - [Untitled]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Add / Modify / Remove an Accessory Application
ClustalW Multiple alignment
BLAST
CAP contig assembly program
ClustalW Example Application
DNADist ---> Neighbor phylogenetic tree
DNADist DNA distance matrix
DNAml DNA Maximum Likelihood program
DNAmlk DNA Maximum Likelihood program with molecular clock
DNAPars DNA parsimony method
FastDNAml DNA maximum likelihood
Fitch -- Fitch-Margoliash and Least-Squares Distance Methods
Kitsch -- Fitch-Margoliash and Least Squares Methods with Evolutionary Clock
NEIGHBOR -- Neighbor-Joining and UPGMA methods
ProML Protein Maximum Likelihood program
Protdist ---> Fitch phylogenetic tree
Protdist ---> Neighbor phylogenetic tree
Protdist protein distance matrix
Protpars protein parsimony method
Courier New
11
Mode: Select / Slide
Start ruler at: 1
Scroll speed slow fast
SEQ01
SEQ02
SEQ03
SEQ04
SEQ05
SEQ06
SEQ07
SEQ08
SEQ09
SEQ10
10
60
70
80
90
100
110
120
130
投影片 3 / 3
"Office 佈景主題"
中文 (台灣)
71%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:38

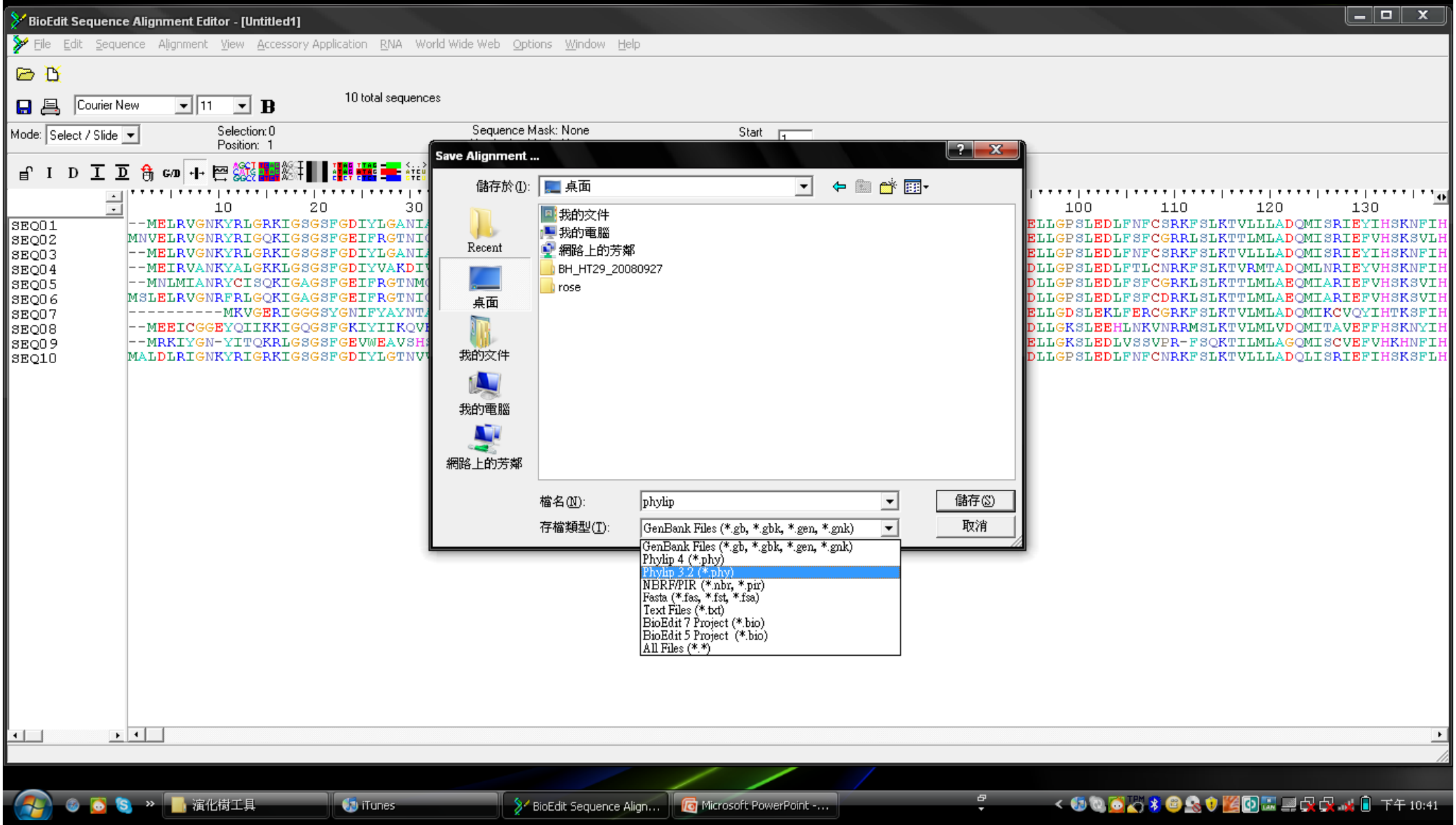

BioEdit Sequence Alignment Editor - [Untitled]
File Edit Sequence Alignment
Courier New
Mode: Select / Slide
ClustalW Options
ClustalW Multiple alignment
Reference:
Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994)
CLUSTAL W: improving the sensitivity of progressive multiple
sequence alignment through sequence weighting, position specific
gap penalties and weight matrix choice.
Nucleic Acids Research, submitted, June 1994.
Full Multiple alignment
Calculate NJ Tree
FAST algorithm for guide tree
Bootstrap NJ Tree
Number of bootstraps: 1000
Gap penalties: Blank=default
Pairwise alignments
Gap open
Gap extend
Multiple alignment
Gap open
Gap extend
Other Parameters:
Note: enter additional parameters as a single line.
Output Clustal format with Clustal consensus sequence generation
Additional Parameters for ClustalW:
General settings:*
/QUICKTREE :use FAST algorithm for the alignment guide tree
/NEWTREE= :file for new guide tree
/USETREE= :file for old guide tree
/NEGATIVE :protein alignment with negative values in matrix
Run ClustalW
View ClustalW Doc
Cancel
SEQ01
SEQ02
SEQ03
SEQ04
SEQ05
SEQ06
SEQ07
SEQ08
SEQ09
SEQ10
fast
70
80
90
100
110
120
130
投影片 4 / 4
"Office 佈景主題"
中文 (台灣)
71%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:38

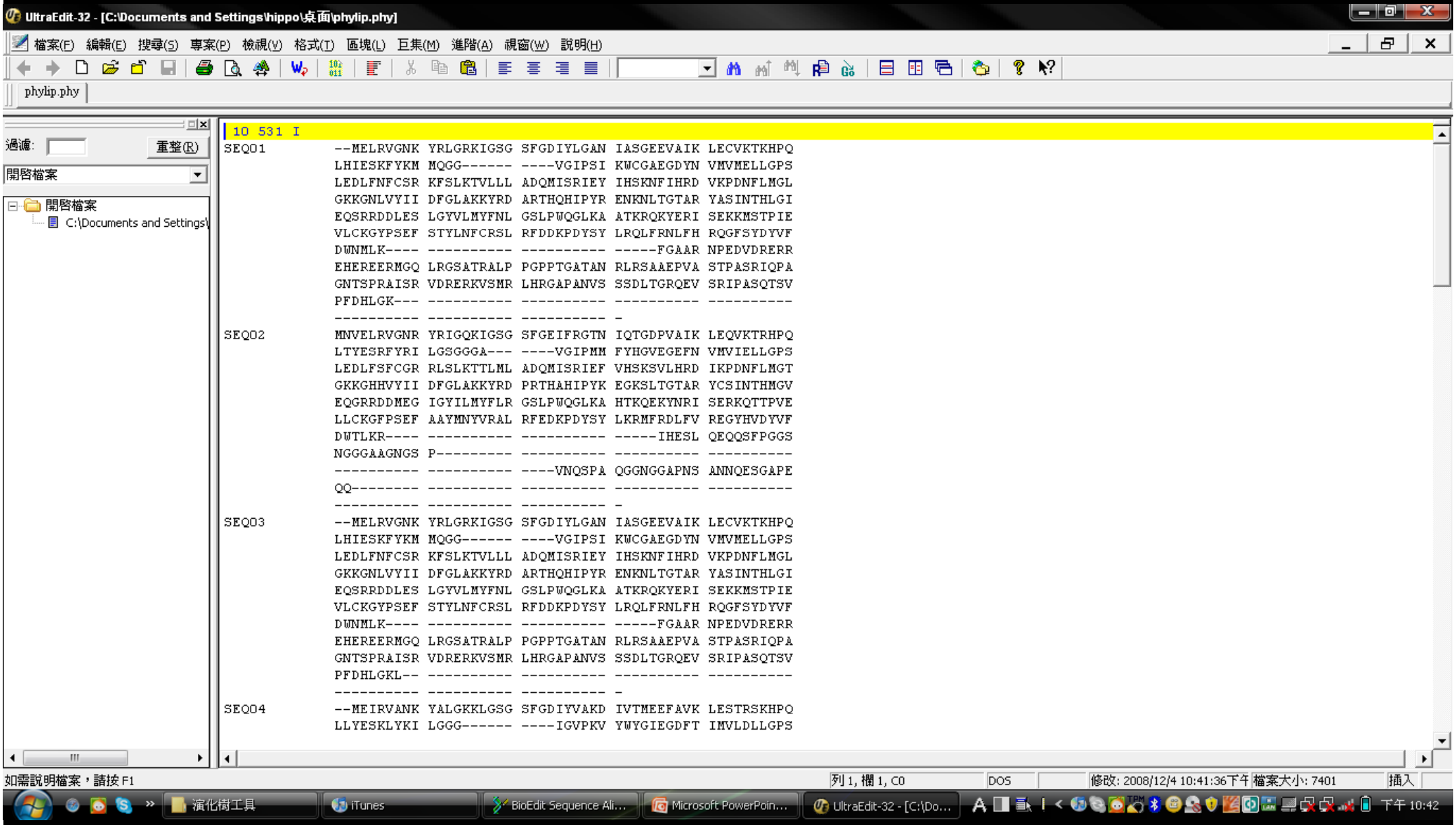

BioEdit Sequence Alignment Editor - [Untitled]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
10 total sequences
Mode: Select / Slide
Selection: null
Position: 5: SEQ05 40
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
SEQ01
SEQ02
SEQ03
SEQ04
SEQ05
SEQ06
SEQ07
SEQ08
SEQ09
SEQ10
BioEdit Sequence Alignment Editor
A log file is being generated.
It will be located in C:\BioEdit\Temp.
The log file for this session will be called BioEdit_clustalw_120408_223925.log.
The temporary input file for this run will be called ~fasta_120408_223925.tmp.
The temporary output file will be called ~align_120408_223925.tmp.
Run parameters will be in a file called clusParm_120408_223925.txt.
All of these files will be located in C:\BioEdit\Temp
Note: These files will not be deleted automatically.
You may want to clean them out of your swap directory periodically.
NOTE THE FOLLOWING IF THIS IS A TRANSLATED NUCLEOTIDE ALIGNMENT!:
Any stops ('*') will be replaced by 'X' for clustal.
They will revert to stops ('*') upon re-toggling the translation.
OK
投影片 6 / 6
"Office 佈景主題"
中文 (台灣)
73%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:39


BioEdit Sequence Alignment Editor - [Untitled]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
C:\BioEdit\apps\clustalw.exe
Sequences (5:7) Aligned. Score: 45
Sequences (5:8) Aligned. Score: 46
Sequences (5:9) Aligned. Score: 43
Sequences (5:10) Aligned. Score: 63
Sequences (6:7) Aligned. Score: 43
Sequences (6:8) Aligned. Score: 45
Sequences (6:9) Aligned. Score: 40
Sequences (6:10) Aligned. Score: 62
Sequences (7:8) Aligned. Score: 38
Sequences (7:9) Aligned. Score: 36
Sequences (7:10) Aligned. Score: 40
Sequences (8:9) Aligned. Score: 34
Sequences (8:10) Aligned. Score: 43
Sequences (9:10) Aligned. Score: 39
Guide tree file created: [C:\BioEdit\Temp\~FASTA~2.dnd]
Start of Multiple Alignment
There are 9 groups
Aligning...
Group 1: Sequences: 2 Score:6443
Group 2: Sequences: 3 Score:6424
Group 3: Sequences: 2 Score:9076
Group 4: Sequences: 3 Score:5874
Group 5: Sequences: 4 Score:6107
Group 6: Sequences: 7 Score:5690
Group 7: Sequences: 8 Score:5294
Mode:
fast
70 80 90 100 110 120 130
SEQ0
SEQ1
t_clustalw_120408_223925.log.
ed ~fasta_120408_223925.tmp.
_120408_223925.tmp.
_120408_223925.txt.
emp
ally.
directory periodically.
TED NUCLEOTIDE ALIGNMENT!:
tal.
g the translation.
投影片 7 / 7
"Office 佈景主題"
中文 (台灣)
73%
演化樹工具
iTunes
BioEdit Sequence Ali...
Microsoft PowerPoin...
C:\BioEdit\apps\clus...
下午 10:39

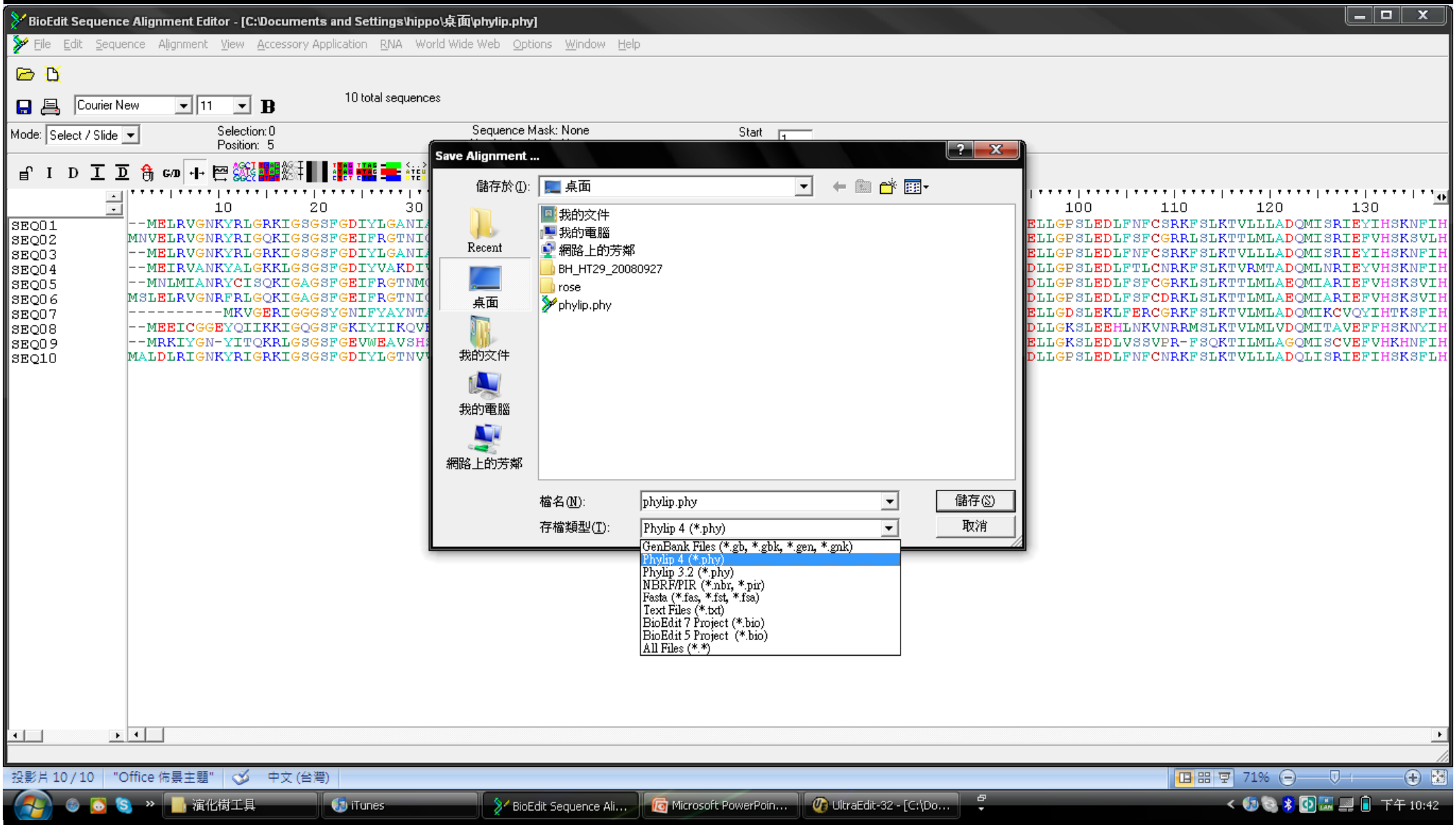

BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
10 total sequences
Mode: Select / Slide
Selection: 0
Position:
Sequence Mask: None
Numbering Mask: None
Start ruler at: 1
Scroll speed slow fast
10 20 30 40 50 60 70 80 90 100 110 120 130
SEQ01 --MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGG----------VGIPSIKWCGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLLADQMISRIEYIHSKNFIH
SEQ02 MNVELRVGNRYRIGQKIGSGSFGEIFRGTNIQTGDPVAIKLEQVKTRHPQLTYESRFYRILGSGGGA-------VGIPMMFYHGVEGEFNVMVIELLGPSLEDLFSFCGRRLSLKTTLMLADQMISRIEFVHSKSVLH
SEQ03 --MELRVGNKYRLGRKIGSGSFGDIYLGANIASGEEVAIKLECVKTKHPQLHIESKFYKMMQGG----------VGIPSIKWCGAEGDYNVMVMELLGPSLEDLFNFCSRKFSLKTVLLLADQMISRIEYIHSKNFIH
SEQ04 --MEIRVANKYALGKKLGSGSFGDIYVAKDIVTMEEFAVKLESTRSKHPQLLYESKLYKILGGG----------IGVPKVYWYGIEGDFTIMVLDLLGPSLEDLFTLCNRKFSLKTVRMTADQMLNRIEYVHSKNFIH
SEQ05 --MNLMIANRYCISQKIGAGSFGEIFRGTNMQTGETVAIKLEQAKTRHPQLAFEARFYRILNAGGGV-------VGIPNILFYGVEGEFNVMVMDLLGPSLEDLFSFCGRKLSLKTTLMLAEQMIARIEFVHSKSVIH
SEQ06 MSLELRVGNRFRLGQKIGAGSFGEIFRGTNIQTGETVAIKLEQAKTRHPQLALEARFYRILNAGGGV-------VGIPNILFYGVEGEFNVMVMDLLGPSLEDLFSFCDRKLSLKTTLMLAEQMIARIEFVHSKSVIH
SEQ07 ----------MKVGERIGGGSYGNIFYAYNTANKKELALKIESEKTKRSQIFNEYRALKCLAGY---------VGIPKVYFETCYGNQNAFTMELLGDSLEKLFERCGRKFSLKTVLMLADQMIKCVQYIHTKSFIH
SEQ08 --MEEICGGEYQIIKKIGQGSFGKIYIIKQVKTGLLFAAKLENSDAPIPQLLFESRLYQIMSGS----------TNVPRLHAHSFDSRYNTIVIDLLGKSLEEHLNKVNRRMSLKTVLMLVDQMITAVEFFHSKNYIH
SEQ09 --MRKIYGN-YITQKRLGSGSFGEVWEAVSHSTGQKVALKLEPRNSSVPQLFFEAKLYSMFQASKSTNNSVEPCNNIPVVYATGQTETTNYMAMELLGKSLEDLVSSVPR-FSQKTILMLAGQMISCVEFVHKHNFIH
SEQ10 MALDLRIGNKYRIGRKIGSGSFGDIYLGTNVVSGEEVAIKLESTRAKHPQLEYEYRVYRILSGG----------VGIPFVRWFGVECDYNAMVMDLLGPSLEDLFNFCNRKFSLKTVLLLADQLISRIEFIHSKSFLH
投影片 8 / 8
"Office 佈景主題"
中文 (台灣)
73%
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:39

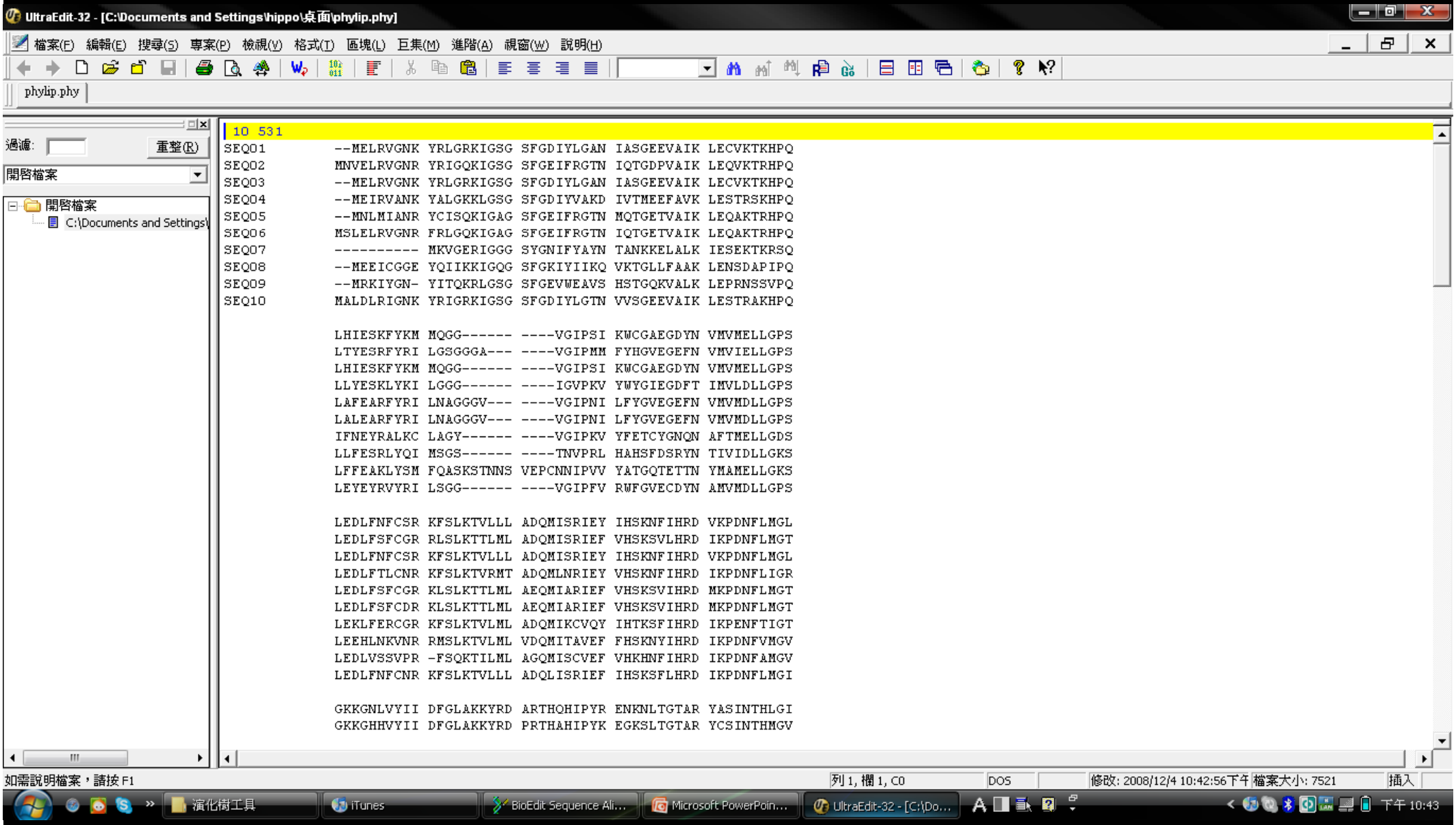
Phylip 4.0

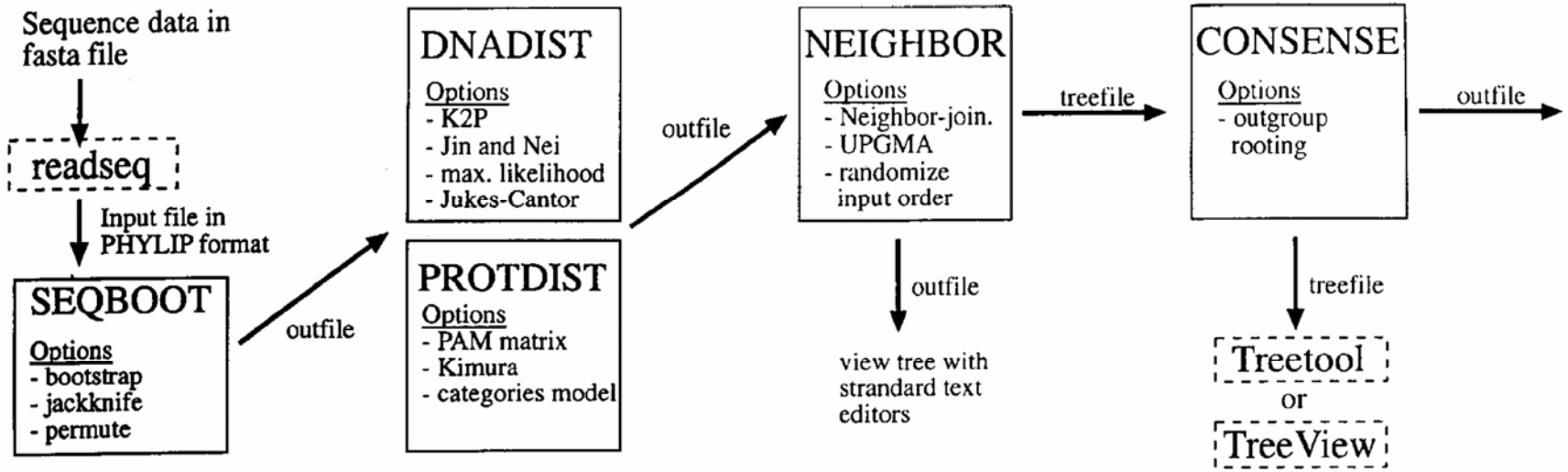

BioEdit Sequence Alignment Editor - [Untitled1]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
10 total sequences
Mode: Select / Slide
Selection: 0
Position: 1
Sequence Mask: None
Start
SEQ01
SEQ02
SEQ03
SEQ04
SEQ05
SEQ06
SEQ07
SEQ08
SEQ09
SEQ10
Save Alignment ...
儲存於(I): 桌面
Recent
桌面
我的文件
我的電腦
網路上的芳鄰
我的文件
我的電腦
BH_HT29_20080927
rose
檔名(N): phylip
存檔類型(T): GenBank Files (*.gb, *.gbk, *.gen, *.gnk)
儲存(S)
取消
GenBank Files (*.gb, *.gbk, *.gen, *.gnk)
Phylip 4 (*.phy)
Phylip 3.2 (*.phy)
NBRF/PIR (*.nbr, *.pir)
Fasta (*.fas, *.fst, *.fsa)
Text Files (*.txt)
BioEdit 7 Project (*.bio)
BioEdit 5 Project (*.bio)
All Files (*.*)
演化樹工具
iTunes
BioEdit Sequence Align...
Microsoft PowerPoint -...
下午 10:41


UltraEdit-32 - [C:\Documents and Settings\hippo\桌面\phylip.phy]
檔案(F) 編輯(E) 搜尋(S) 專案(P) 檢視(V) 格式(T) 區塊(L) 巨集(M) 進階(A) 視窗(W) 說明(H)
phylip.phy
過濾:
重整(R)
開啓檔案
開啓檔案
C:\Documents and Settings\
10 531 I
SEQ01 --MELRVGNK YRLGRKIGSG SFGDIYLGAN IASGEEVAIK LECVKTKHPQ
LHIESKFYKM MQGG------ ----VGIPSI KWCGAEGDYN VMVMELLGPS
LEDLFNFCSR KFSLKTVLLL ADQMISRIEY IHSKNFIHRD VKPDNFLMGL
GKKGNLVYII DFGLAKKYRD ARTHQHIPYR ENKNLTGTAR YASINTHLGI
EQSRRDDLES LGYVLMYFNL GSLPWQGLKA ATKRQKYERI SEKKMSTPIE
VLCKGYPSEF STYLNFCRSL RFDDKPDYSY LRQLFRNLFH RQGFSYDYVF
DWNMLK---- ---------- ---------- -----FGAAR NPEDVDRERR
EHEREERMGQ LRGSATRALP PGPPTGATAN RLRSAAEPVA STPASRIQPA
GNTSPRAISR VDRERKVSMR LHRGAPANVS SSDLTGRQEV SRIPASQTSV
PFDHLGK--- ---------- ---------- ---------- ----------
---------- ---------- ---------- -
SEQ02 MNVELRVGNR YRIGQKIGSG SFGEIFRGTN IQTGDPVAIK LEQVKTRHPQ
LTYESRFYRI LGSGGGA--- ----VGIPMM FYHGVEGEFN VMVIELLGPS
LEDLFSFCGR RLSLKTTLML ADQMISRIEF VHSKSVLHRD IKPDNFLMGT
GKKGHHVYII DFGLAKKYRD PRTHAHIPYK EGKSLTGTAR YCSINTHMGV
EQGRRDDMEG IGYILMYFLR GSLPWQGLKA HTKQEKYNRI SERKQTTPVE
LLCKGFPSEF AAYMNYVRAL RFEDKPDYSY LKRMFRDLFV REGYHVDYVF
DWTLKR---- ---------- ---------- -----IHESL QEQQSFPGGS
NGGGAAGNGS P--------- ---------- ---------- ----------
---------- ---------- ----VNQSPA QGGNGGAPNS ANNQESGAPE
QQ-------- ---------- ---------- ---------- ----------
---------- ---------- ---------- -
SEQ03 --MELRVGNK YRLGRKIGSG SFGDIYLGAN IASGEEVAIK LECVKTKHPQ
LHIESKFYKM MQGG------ ----VGIPSI KWCGAEGDYN VMVMELLGPS
LEDLFNFCSR KFSLKTVLLL ADQMISRIEY IHSKNFIHRD VKPDNFLMGL
GKKGNLVYII DFGLAKKYRD ARTHQHIPYR ENKNLTGTAR YASINTHLGI
EQSRRDDLES LGYVLMYFNL GSLPWQGLKA ATKRQKYERI SEKKMSTPIE
VLCKGYPSEF STYLNFCRSL RFDDKPDYSY LRQLFRNLFH RQGFSYDYVF
DWNMLK---- ---------- ---------- -----FGAAR NPEDVDRERR
EHEREERMGQ LRGSATRALP PGPPTGATAN RLRSAAEPVA STPASRIQPA
GNTSPRAISR VDRERKVSMR LHRGAPANVS SSDLTGRQEV SRIPASQTSV
PFDHLGKL-- ---------- ---------- ---------- ----------
---------- ---------- ---------- -
SEQ04 --MEIRVANK YALGKKLGSG SFGDIYVAKD IVTMEEFAVK LESTRSKHPQ
LLYESKLYKI LGGG------ ----IGVPKV YWYGIEGDFT IMVLDLLGPS
如需說明檔案，請按 F1
列 1, 欄 1, C0
DOS
修改: 2008/12/4 10:41:36下午
檔案大小: 7401
插入
演化樹工具
iTunes
BioEdit Sequence Ali...
Microsoft PowerPoin...
UltraEdit-32 - [C:\Do...
下午 10:42

Phylip 3.2


BioEdit Sequence Alignment Editor - [C:\Documents and Settings\hippo\桌面\phylip.phy]
File Edit Sequence Alignment View Accessory Application RNA World Wide Web Options Window Help
Courier New
11
10 total sequences
Mode: Select / Slide
Selection: 0
Position: 5
Sequence Mask: None
Start
SEQ01
SEQ02
SEQ03
SEQ04
SEQ05
SEQ06
SEQ07
SEQ08
SEQ09
SEQ10
Save Alignment ...
儲存於(I): 桌面
Recent
桌面
我的文件
我的電腦
網路上的芳鄰
我的文件
我的電腦
網路上的芳鄰
BH_HT29_20080927
rose
phylip.phy
檔名(N): phylip.phy
存檔類型(T): Phylip 4 (*.phy)
儲存(S)
取消
GenBank Files (*.gb, *.gbk, *.gen, *.gnk)
Phylip 4 (*.phy)
Phylip 3.2 (*.phy)
NBRF/PIR (*.nbr, *.pir)
Fasta (*.fas, *.fst, *.fsa)
Text Files (*.txt)
BioEdit 7 Project (*.bio)
BioEdit 5 Project (*.bio)
All Files (*.*)
投影片 10 / 10
"Office 佈景主題"
中文 (台灣)
71%
演化樹工具
iTunes
BioEdit Sequence Ali...
Microsoft PowerPoin...
UltraEdit-32 - [C:\Do...
下午 10:42


UltraEdit-32 - [C:\Documents and Settings\hippo\桌面\phylip.phy]
檔案(F) 編輯(E) 搜尋(S) 專案(P) 檢視(V) 格式(T) 區塊(L) 巨集(M) 進階(A) 視窗(W) 說明(H)
phylip.phy
過濾: 重整(R)
開啓檔案
開啓檔案
C:\Documents and Settings\
10 531
SEQ01 --MELRVGNK YRLGRKIGSG SFGDIYLGAN IASGEEVAIK LECVKTKHPQ
SEQ02 MNVELRVGNR YRIGQKIGSG SFGEIFRGTN IQTGDPVAIK LEQVKTRHPQ
SEQ03 --MELRVGNK YRLGRKIGSG SFGDIYLGAN IASGEEVAIK LECVKTKHPQ
SEQ04 --MEIRVANK YALGKKLGSG SFGDIYVAKD IVTMEEFAVK LESTRSKHPQ
SEQ05 --MNLMIANR YCISQKIGAG SFGEIFRGTN MQTGETVAIK LEQAKTRHPQ
SEQ06 MSLELRVGNR FRLGQKIGAG SFGEIFRGTN IQTGETVAIK LEQAKTRHPQ
SEQ07 ---------- MKVGERIGGG SYGNIFYAYN TANKKELALK IESEKTKRSQ
SEQ08 --MEEICGGE YQIIKKIGQG SFGKIYIIKQ VKTGLLFAAK LENSDAPIPQ
SEQ09 --MRKIYGN- YITQKRLGSG SFGEVWEAVS HSTGQKVALK LEPRNSSVPQ
SEQ10 MALDLRIGNK YRIGRKIGSG SFGDIYLGTN VVSGEEVAIK LESTRAKHPQ
LHIESKFYKM MQGG------ ----VGIPSI KWCGAEGDYN VMVMELLGPS
LTYESRFYRI LGSGGGA--- ----VGIPMM FYHGVEGEFN VMVIELLGPS
LHIESKFYKM MQGG------ ----VGIPSI KWCGAEGDYN VMVMELLGPS
LLYESKLYKI LGGG------ ----IGVPKV YWYGIEGDFT IMVLDLLGPS
LAFEARFYRI LNAGGGV--- ----VGIPNI LFYGVEGEFN VMVMDLLGPS
LALEARFYRI LNAGGGV--- ----VGIPNI LFYGVEGEFN VMVMDLLGPS
IFNEYRALKC LAGY------ ----VGIPKV YFETCYGNQN AFTMELLGDS
LLFESRLYQI MSGS------ ----TNVPRL HAHSFDSRYN TIVIDLLGKS
LFFEAKLYSM FQASKSTNNS VEPCNNIPVV YATGQTETTN YMAMELLGKS
LEYEYRVYRI LSGG------ ----VGIPFV RWFGVECDYN AMVMDLLGPS
LEDLFNFCSR KFSLKTVLLL ADQMISRIEY IHSKNFIHRD VKPDNFLMGL
LEDLFSFCGR RLSLKTTLML ADQMISRIEF VHSKSVLHRD IKPDNFLMGT
LEDLFNFCSR KFSLKTVLLL ADQMISRIEY IHSKNFIHRD VKPDNFLMGL
LEDLFTLCNR KFSLKTVRMT ADQMLNRIEY VHSKNFIHRD IKPDNFLIGR
LEDLFSFCGR KLSLKTTLML AEQMIARIEF VHSKSVIHRD MKPDNFLMGT
LEDLFSFCDR KLSLKTTLML AEQMIARIEF VHSKSVIHRD MKPDNFLMGT
LEKLFERCGR KFSLKTVLML ADQMIKCVQY IHTKSFIHRD IKPENFTIGT
LEEHLNKVNR RMSLKTVLML VDQMITAVEF FHSKNYIHRD IKPDNFVMGV
LEDLVSSVPR -FSQKTILML AGQMISCVEF VHKHNFIHRD IKPDNFAMGV
LEDLFNFCNR KFSLKTVLLL ADQLISRIEF IHSKSFLHRD IKPDNFLMGI
GKKGNLVYII DFGLAKKYRD ARTHQHIPYR ENKNLTGTAR YASINTHLGI
GKKGHHVYII DFGLAKKYRD PRTHAHIPYK EGKSLTGTAR YCSINTHMGV
如需說明檔案，請按 F1
列 1, 欄 1, C0
DOS
修改: 2008/12/4 10:42:56下午
檔案大小: 7521
插入
演化樹工具
iTunes
BioEdit Sequence Ali...
Microsoft PowerPoin...
UltraEdit-32 - [C:\Do...
下午 10:43



PHYLIP input file

```
 5 60
human   TKYHSDDYST GGSVASAVNW AG..
mouse   TKYHSDEYST GGSVAGAVNW AG..
frog    TKYHSDDYSA GGSVASSVNW SG..
yeast   CQFHTDEYSG GGSMEGAVNY AG..
bact.   ELVHEPDYVG GTILTAACNT AG..
```

SEQBOOT output

```
    5    60
human   TTTKKYDDDY YYTGGGSASA VV..
mouse   TTTKKYDEEY YYTGGGSAGA VV..
frog    TTTKKYDDDY YYAGGGSASS VV..
yeast   CCCQQFDEEY YYGGGGSEGA VV..
bact.   EEELLVPDDY YYGGGTITAA CC..
    5    60
human   TKKYSSSDYS GGSAAVNWWW AA..
mouse   TKKYSSSEYS GGGAAVNWWW AA..
frog    TKKYSSSDYS GGSSSVNWWW SS..
yeast   CQQFTTTEYS GGGAAVNYYY AA..
bact.   ELLVEEEDYV GGAAACNTTT AA..
    5    60
human   TYYYHHSDDD DYYSTTASAA VV..
mouse   TYYYHHSDDD EYYSTTAGAA VV..
  .               .           .
  .               .           .
```

PROTDIST output

```
    5
human   0.00000  0.00982  0.10084  0.24911  1.01684
mouse   0.00982  0.00000  0.13747  0.21005  1.05809
frog    0.10084  0.13747  0.00000  0.34655  1.19283
yeast   0.24911  0.21005  0.34655  0.00000  0.97720
bact.   1.01684  1.05809  1.19283  0.97720  0.00000
    5
human   0.00000  0.00000  0.08682  0.41532  1.13122
mouse   0.00000  0.00000  0.08682  0.41532  1.13122
  .        .        .        .        .        .
  .        .        .        .        .        .
```

NEIGHBOR/CONSENSE outfile

```
          +---------mouse
      +-93.0
      !   !     +----frog
+-94.0    +-72.0
!     !         +----human
!     !
!     +--------------yeast
!
+-------------------bacterium
```

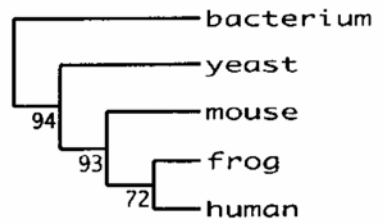
NEIGHBOR treefile

```
(frog:0.09252,((yeast:0.20912,bacterium:1.29
294):0.16052,mouse:-0.00846):0.02131,human:-
0.02075);

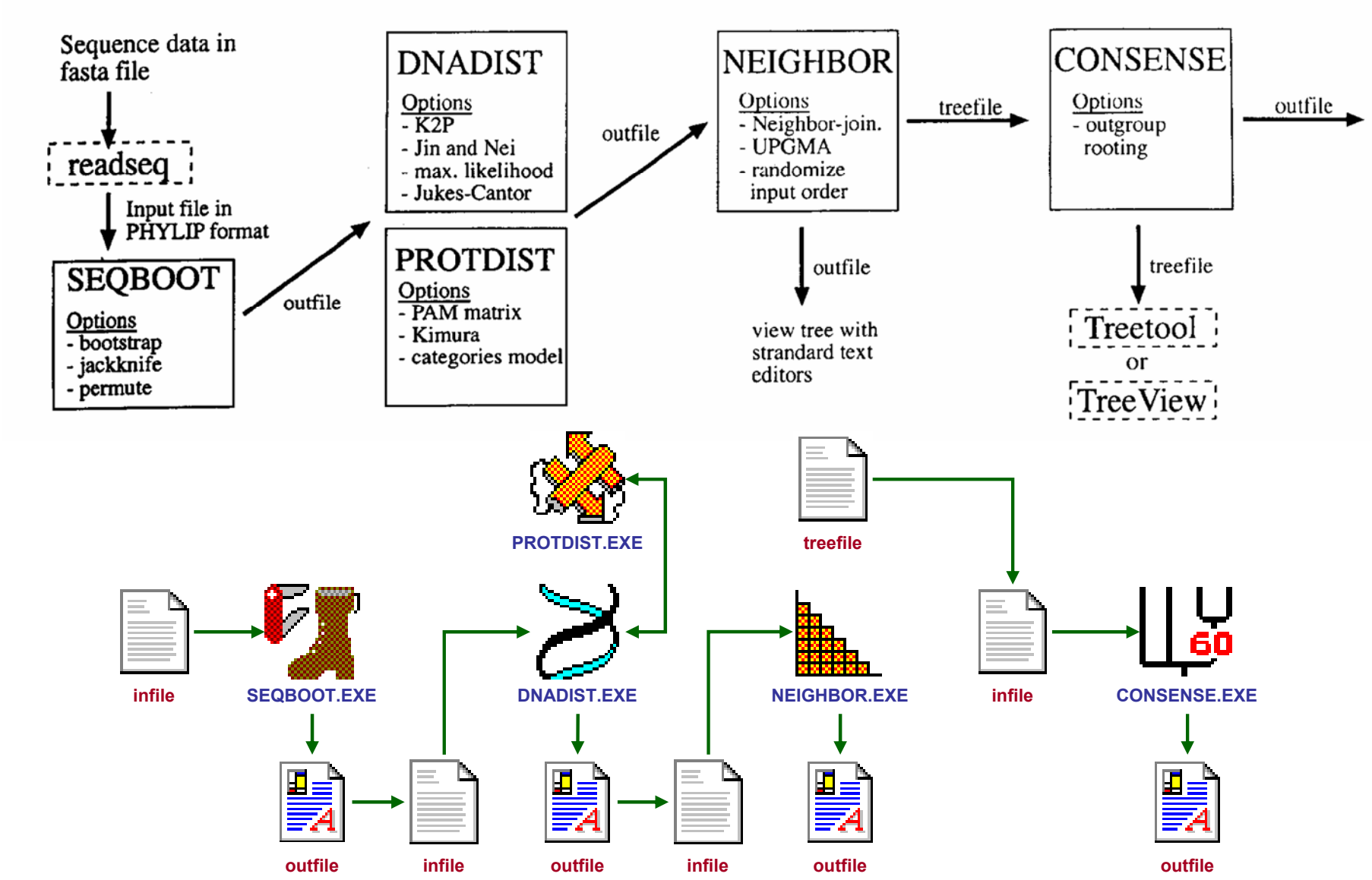
(mouse:-0.00225,frog:0.09130,(human:-0.00225,
(yeast:0.16225,bacterium:0.63517):0.11338):0
.00225);

((frog:0.14362,human:-0.01934):0.03099,(yeast:
0.12204,bacterium:0.80782):0.10822,mouse:-
0.00667);
```

Treeview output



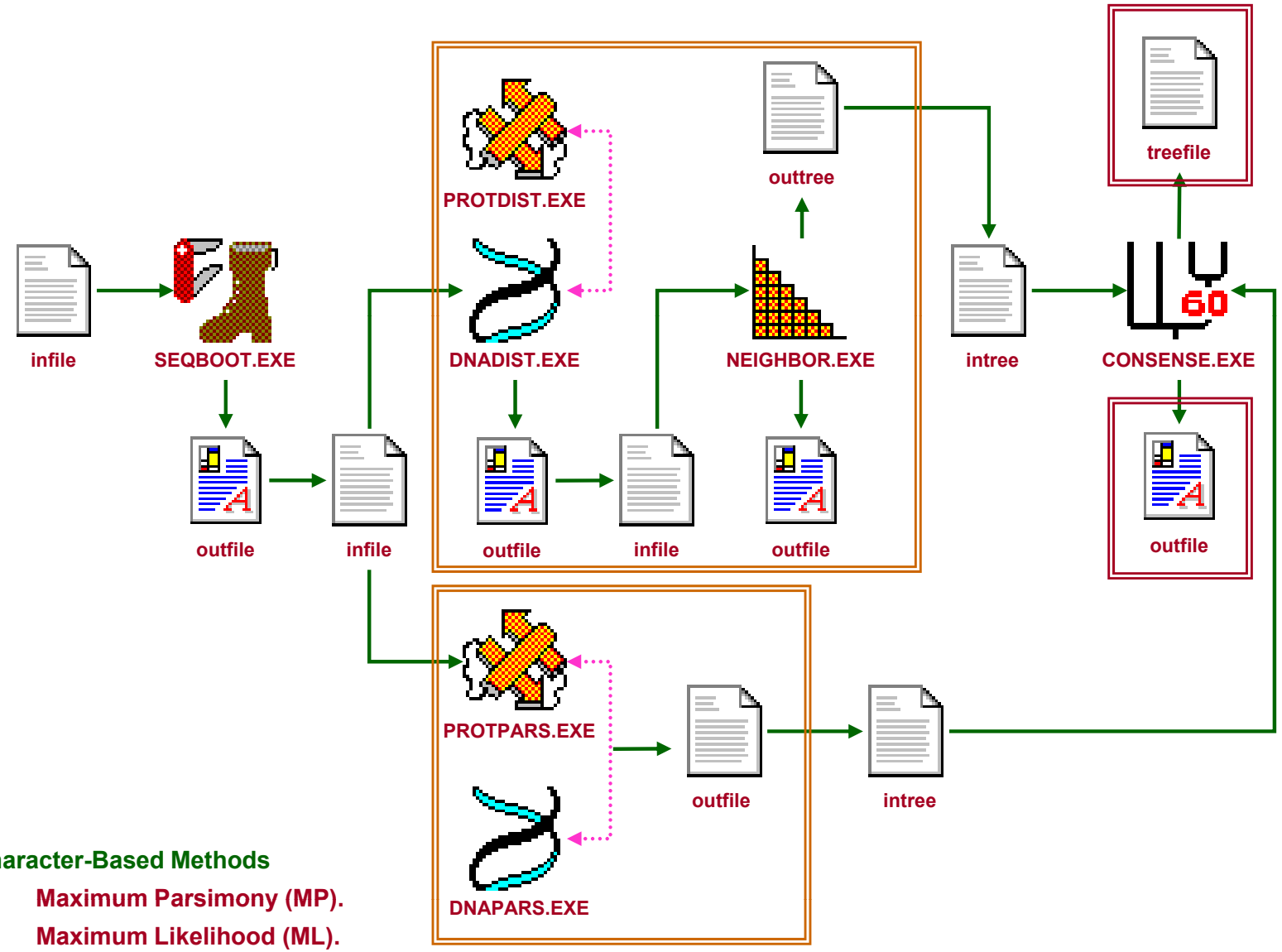
Tree build pipeline



Tree Generation Flowchart

Distance-Based Methods

1. Unweighted Pair Group Method with ArithmeticMean (UPGMA).
2. Neighbor Joining (NJ).
3. Fitch-Margoliash (FM).
4. Minimum Evolution (ME).



Character-Based Methods

1. Maximum Parsimony (MP).
2. Maximum Likelihood (ML).

Sequence alignment and trimming



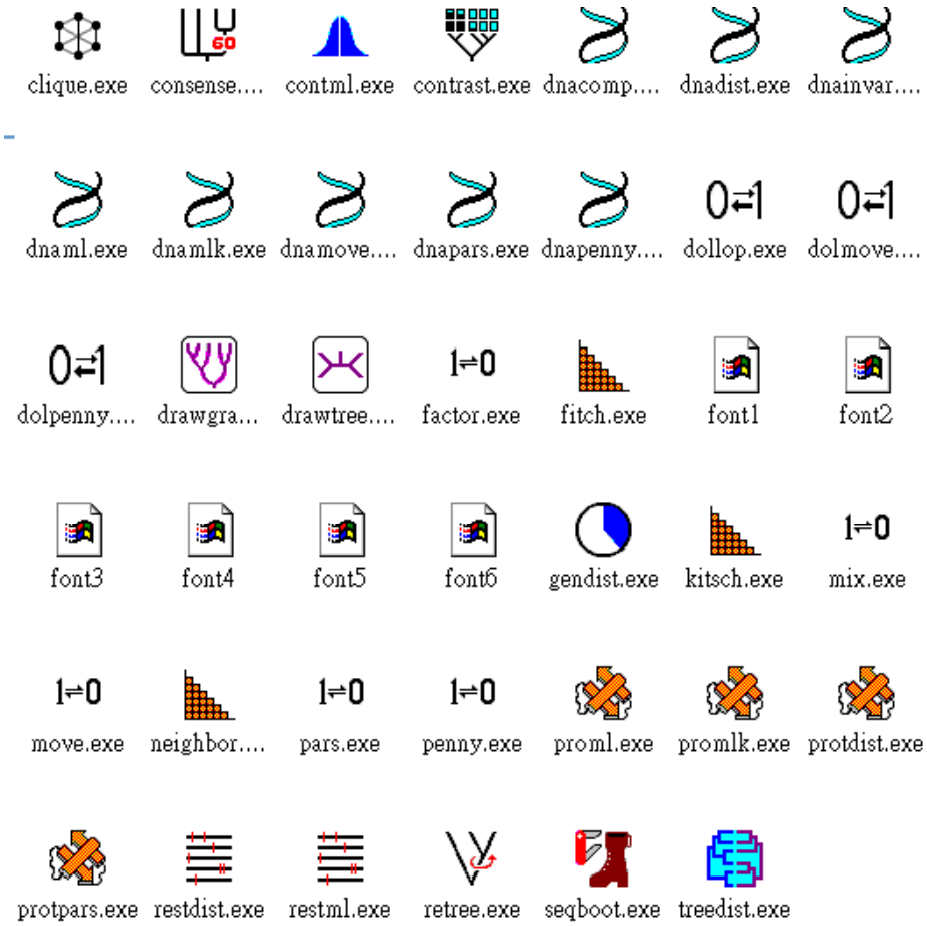
Get Programs

... by type of data

- DNA sequences
- Protein sequences
- Restriction sites
- Distance matrices
- Gene frequencies
- Quantitative characters
- Discrete characters
- tree plotting, consensus trees, tree distances and tree manipulation

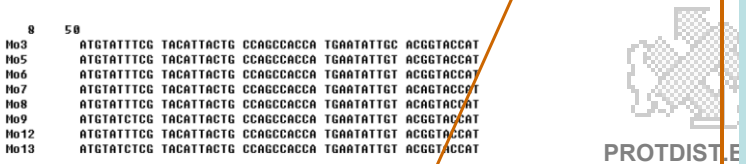
... by type of algorithm

- Heuristic tree search
- Branch-and-bound tree search
- Interactive tree manipulation
- Plotting trees, consenus trees, tree distances
- Converting data, making distances or bootstrap replicates

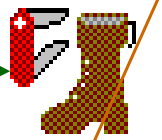


Step 1.1

1. **Bootstraping** 就是從整個序列的城基（氨基酸）中任意選取一半，剩下的一半序列隨機補齊組成一個新的序列。這樣，一個序列就可以變成了許多序列。一個多序列組也就可以變成許多個多序列組。根據某種演算法（最大簡約性法、最大可能性法、除權配對法或鄰位相連法）每個多序列組都可以生成一個進化樹。將生成的許多進化樹進行比較，按照多數規則（majority-rule）我們就會得到一個最“逼真”的進化樹。
2. **Jackknife** 則是另外一種隨機選取序列的方法。它與 Bootstrap 法的區別是不將剩下的一半序列補齊，只生成一個縮短了一半的新序列。
3. **Permute** 是將一個數組中的元素的順序隨機化。



infile



SEQBOOT.EXE

PROTDIST.E

DNADIST.E



outfile



infile

outfile

intree

outfile

outfile

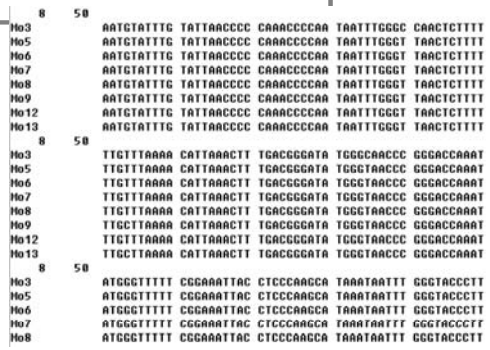
```
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\SEQBOOT.EXE:  can't read infile
Please enter a new filename>D:\1.phy

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

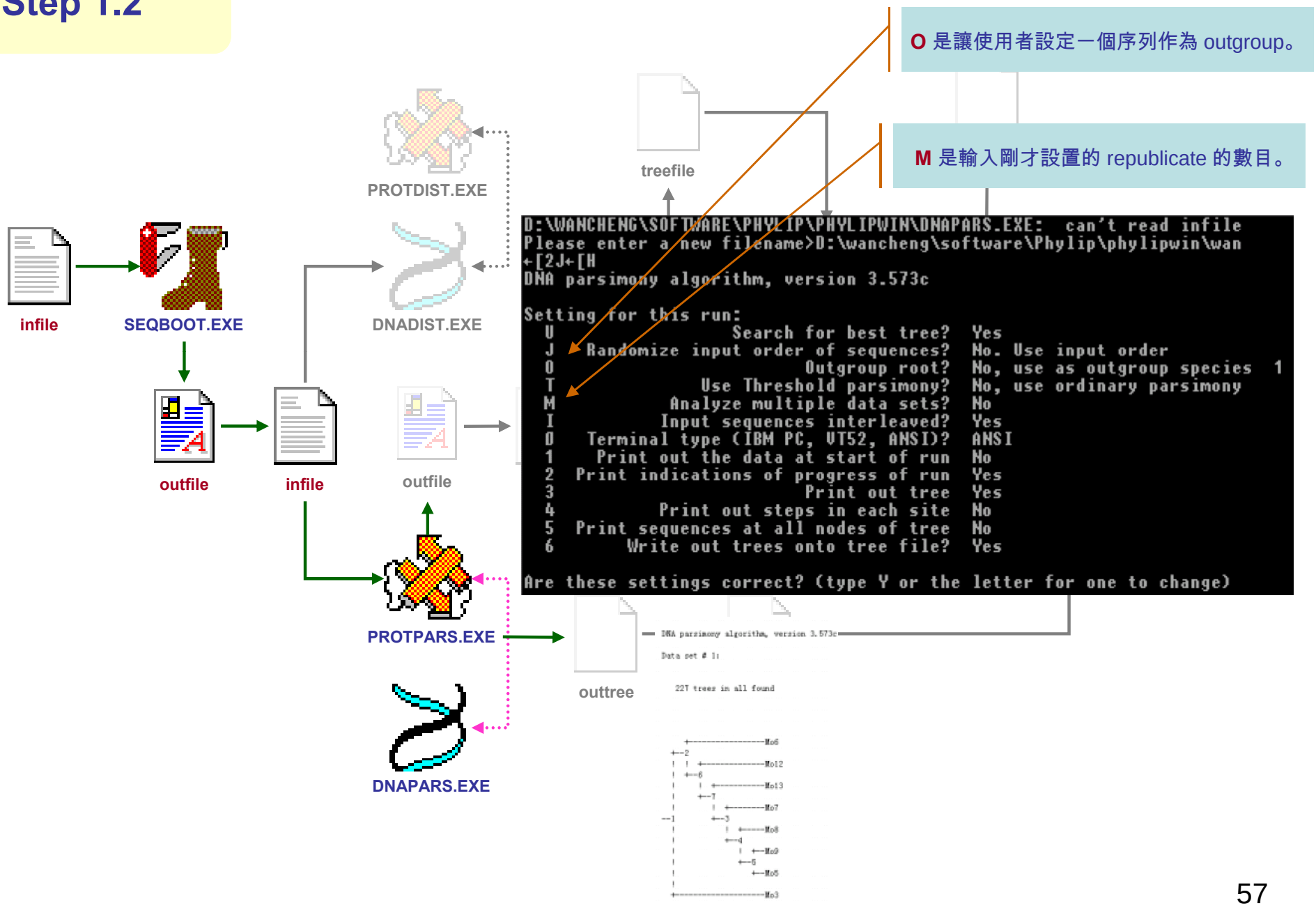
Settings for this run:
  D   Sequence, Morph, Rest., Gene Freqs?  Molecular sequences
  J      Bootstrap, Jackknife, or Permute?  Bootstrap
  R                    How many replicates?  100
  I           Input sequences interleaved?  Yes
  0    Terminal type (IBM PC, VT52, ANSI)?  ANSI
  1     Print out 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)
_
```

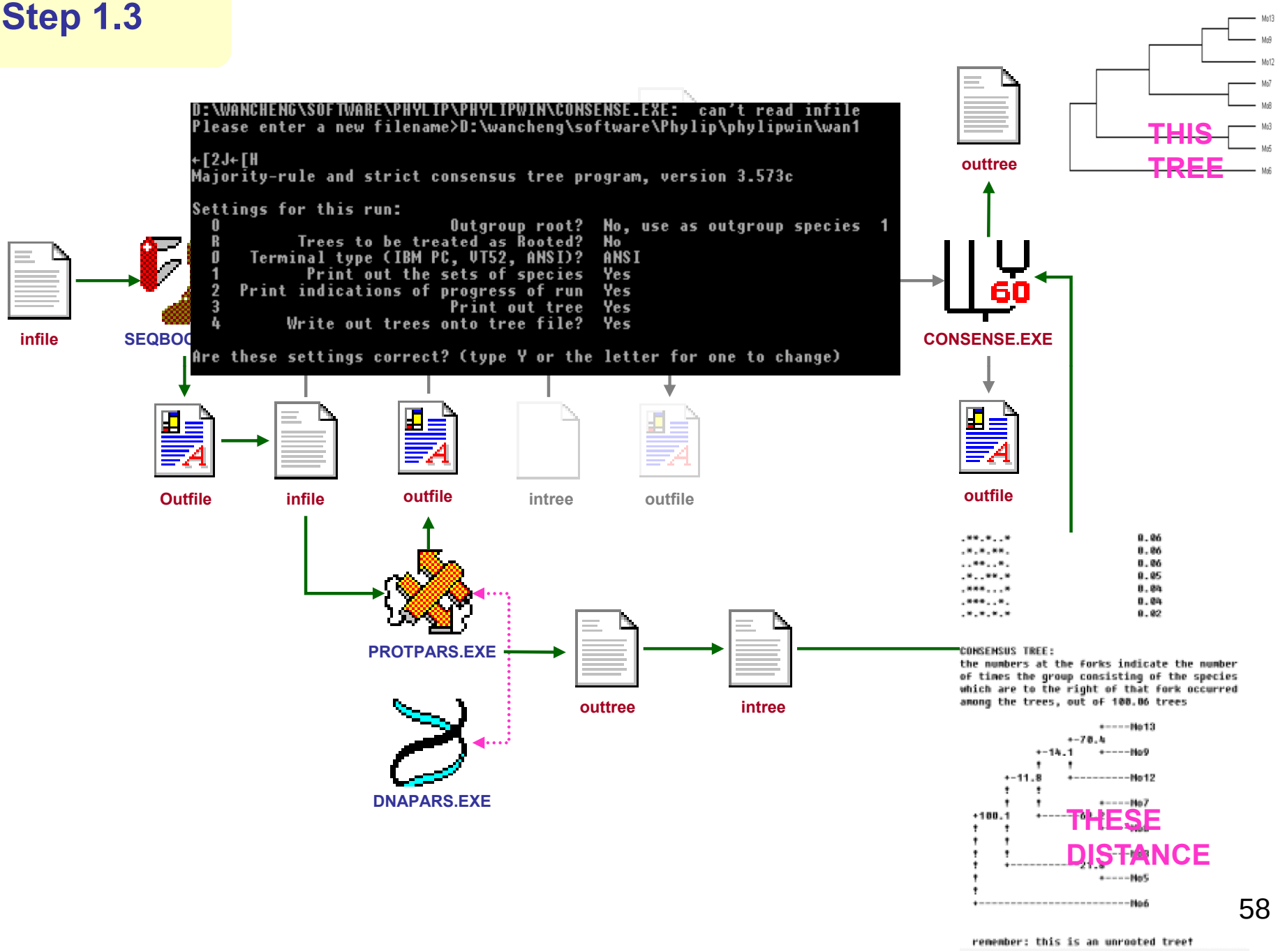
Republicate 就是用 Bootstrap 法生成的一個多序列組。



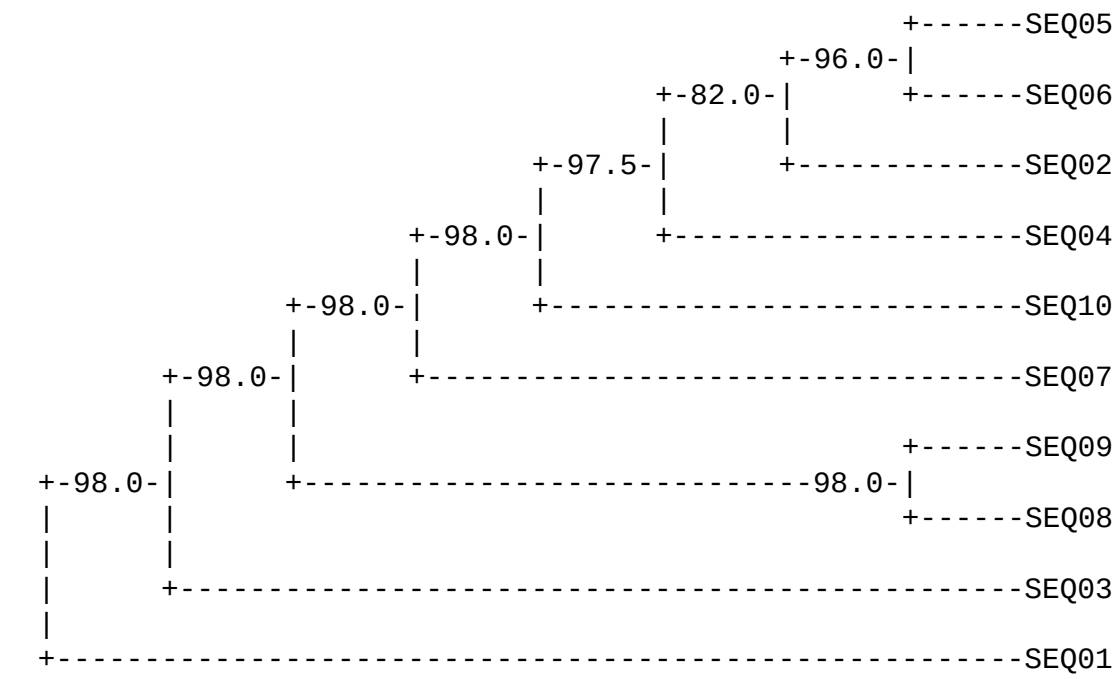
Step 1.2

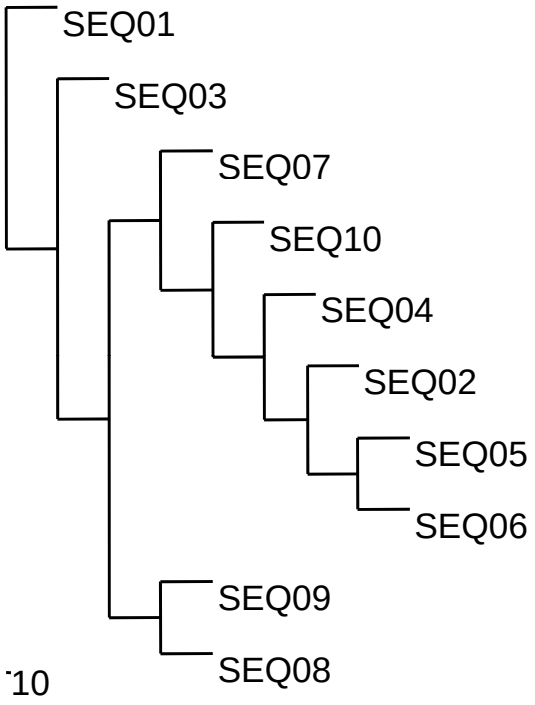

O 是讓使用者設定一個序列作為 outgroup。
M 是輸入剛才設置的 republicate 的數目。
treefile
PROTDIST.EXE
infile
SEQBOOT.EXE
DNADIST.EXE
outfile
infile
outfile
PROTPARS.EXE
outtree
DNAPARS.EXE
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\DNAPARS.EXE: can't read infile
Please enter a new filename>D:\wancheng\software\Phylip\phylipwin\wan
←[2J←[H
DNA parsimony algorithm, version 3.573c
Setting for this run:
U Search for best tree? Yes
J Randomize input order of sequences? No. Use input order
O Outgroup root? No, use as outgroup species 1
T Use Threshold parsimony? No, use ordinary parsimony
M Analyze multiple data sets? No
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
3 Print out tree Yes
4 Print out steps in each site No
5 Print sequences at all nodes of tree No
6 Write out trees onto tree file? Yes
Are these settings correct? (type Y or the letter for one to change)

Step 1.3


infile
SEQBOO
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)
outtree
THIS TREE
CONSENSE.EXE
Outfile
infile
outfile
intree
outfile
outfile
PROTPARS.EXE
DNAPARS.EXE
outtree
intree
CONSENSUS TREE:
the numbers at the forks indicate the number
of times the group consisting of the species
which are to the right of that fork occurred
among the trees, out of 100.00 trees
THESE DISTANCE
remember: this is an unrooted tree!

```
CONSENSUS TREE:
the numbers forks indicate the number
of times the group consisting of the species
which are to the right of that fork occurred
among the trees, out of  98.00 trees
```

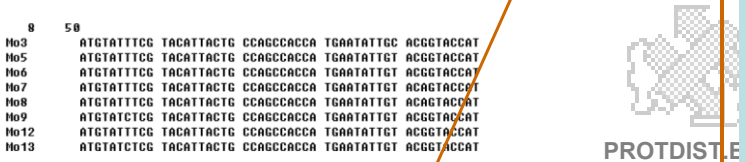


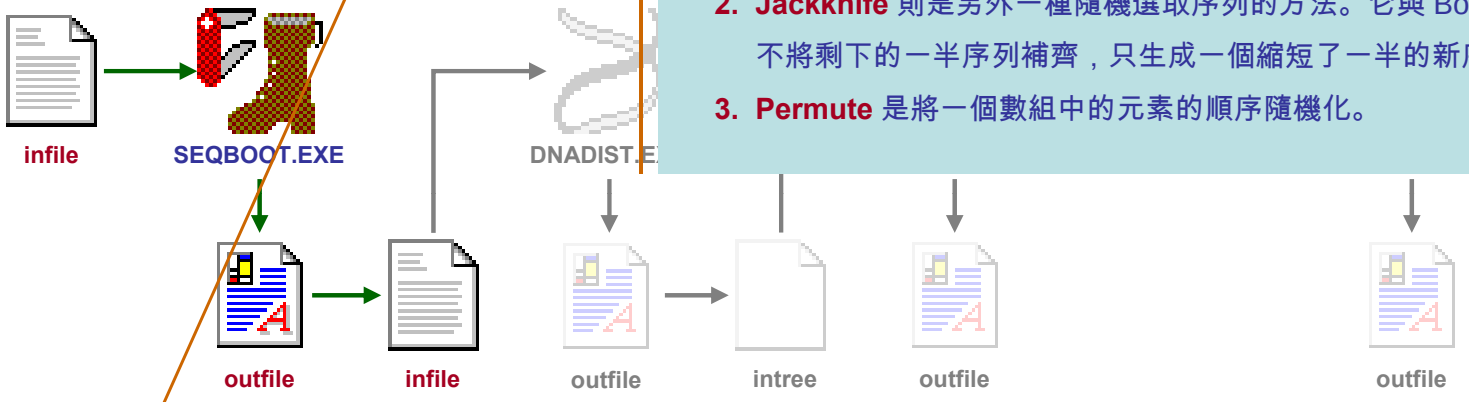


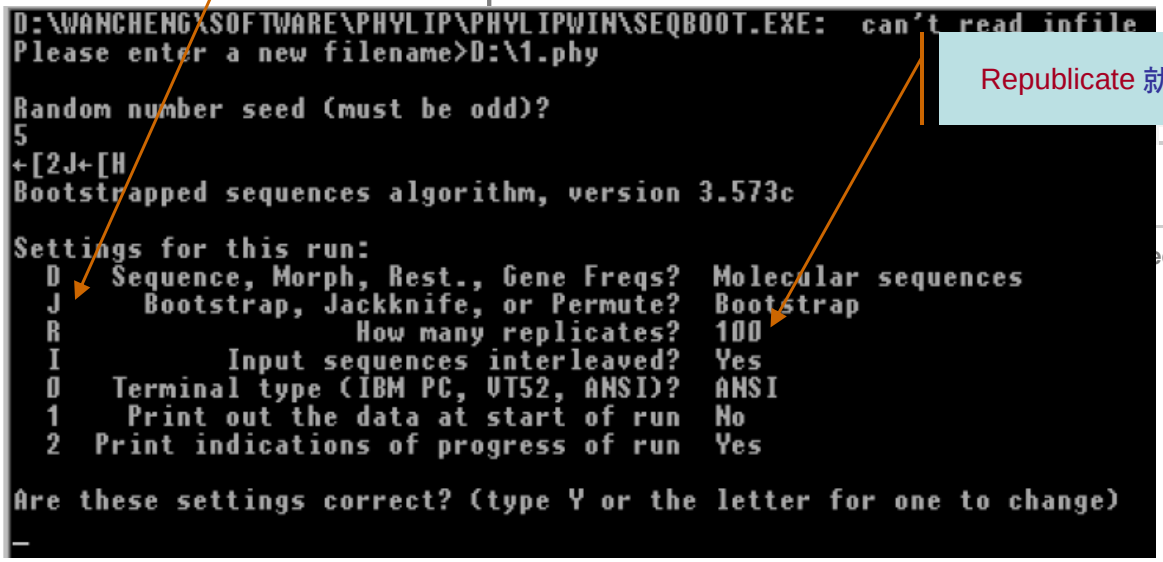
rooted

Step 2.1

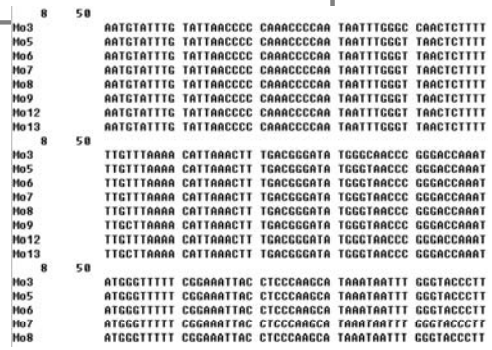
1. Bootstraping 就是從整個序列的城基（氨基酸）中任意選取一半，剩下的一半序列隨機補齊組成一個新的序列。這樣，一個序列就可以變成了許多序列。一個多序列組也就可以變成許多個多序列組。根據某種演算法（最大簡約性法、最大可能性法、除權配對法或鄰位相連法）每個多序列組都可以生成一個進化樹。將生成的許多進化樹進行比較，按照多數規則（majority-rule）我們就會得到一個最“逼真”的進化樹。
2. **Jackknife** 則是另外一種隨機選取序列的方法。它與 Bootstrap 法的區別是不將剩下的一半序列補齊，只生成一個縮短了一半的新序列。
3. **Permute** 是將一個數組中的元素的順序隨機化。



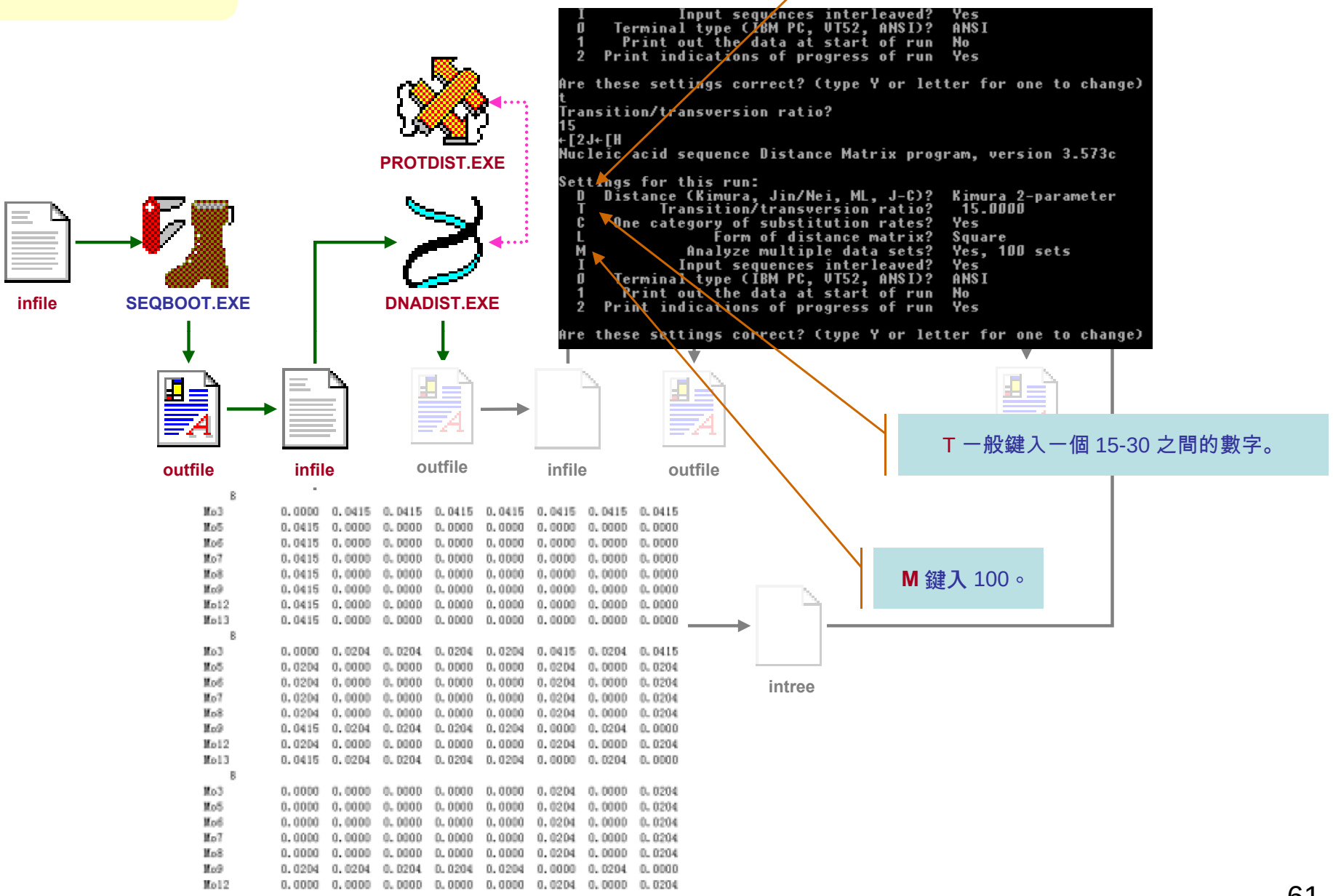




Republicate 就是用 Bootstrap 法生成的一個多序列組。



Step 2.2

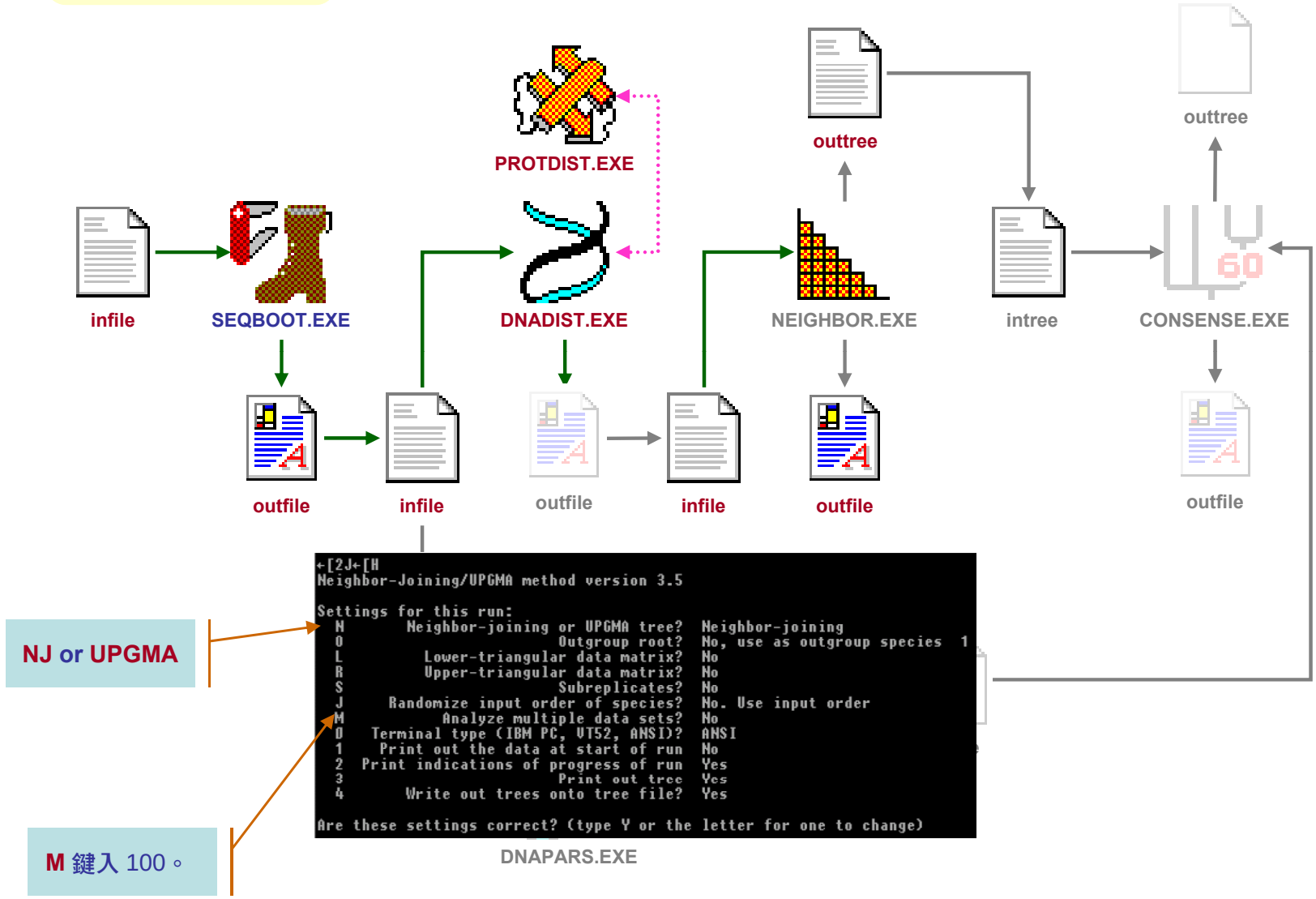


D 有四種距離模式可以選擇，分別是 Kimura 2-parameter、Jin/Nei、Maximum-likelihood 和 Jukes-Cantor。

T 一般鍵入一個 15-30 之間的數字。

M 鍵入 100。

Step 2.3

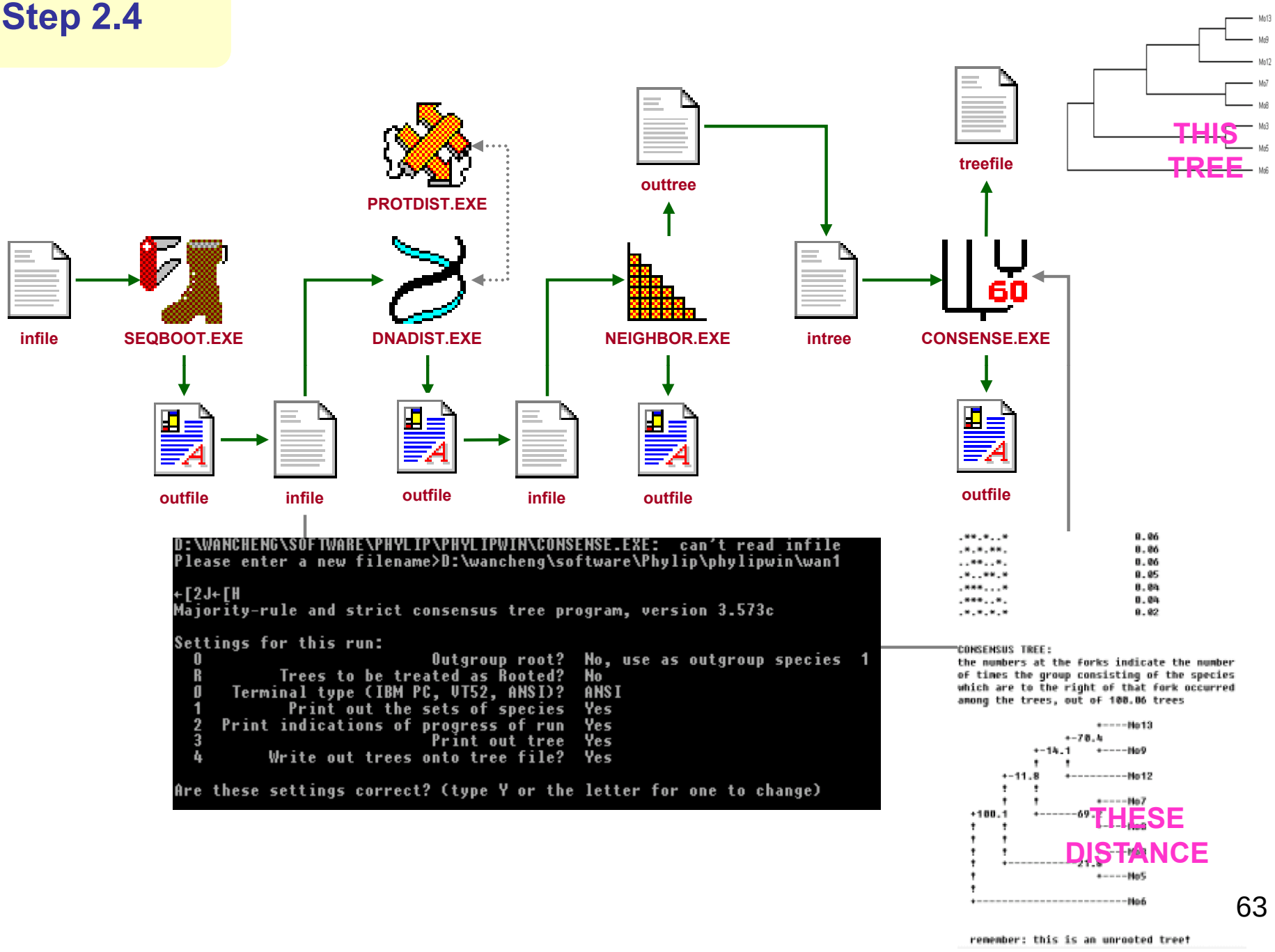



```
←[2J←[H
Neighbor-Joining/UPGMA method version 3.5

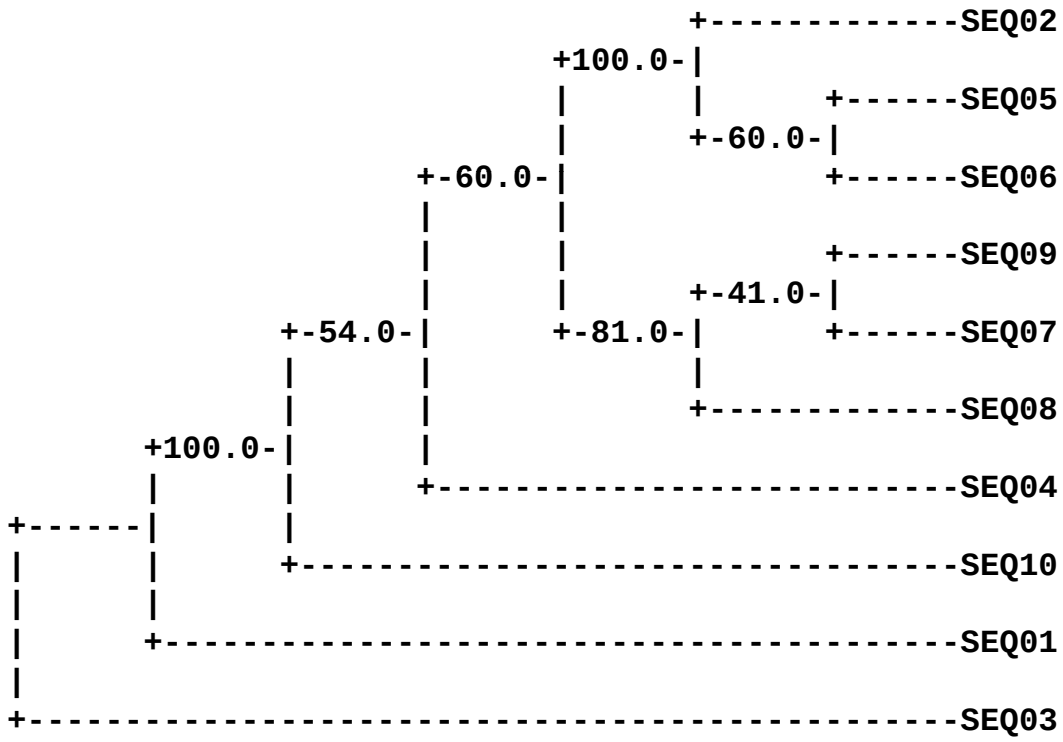
Settings for this run:
  N       Neighbor-joining or UPGMA tree?  Neighbor-joining
  O                      Outgroup root?  No, use as outgroup species  1
  L         Lower-triangular data matrix?  No
  R         Upper-triangular data matrix?  No
  S                      Subreplicates?  No
  J     Randomize 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

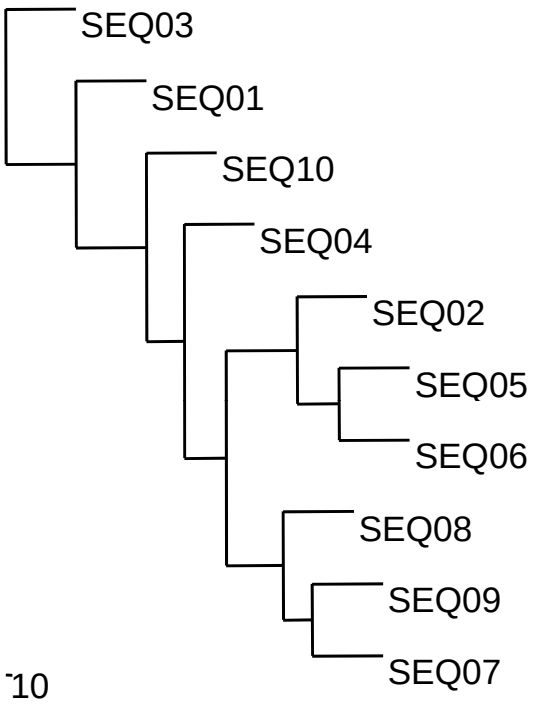
Are these settings correct? (type Y or the letter for one to change)
```

Step 2.4

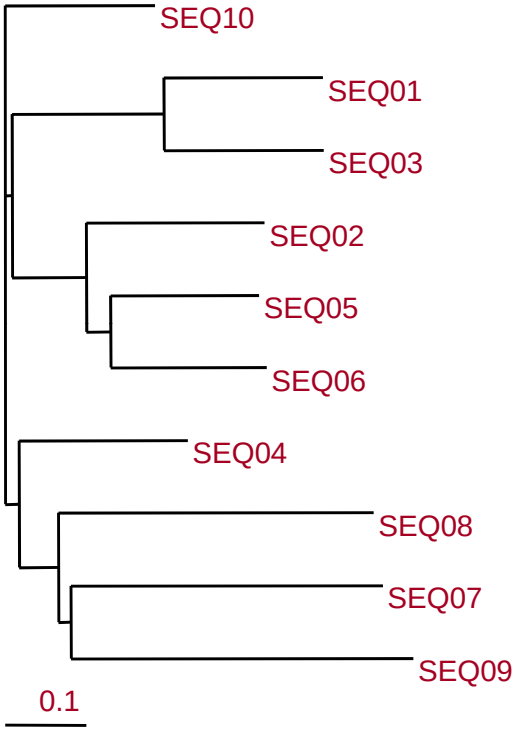

infile
SEQBOOT.EXE
PROTDIST.EXE
DNADIST.EXE
NEIGHBOR.EXE
outtree
intree
CONSENSE.EXE
treefile
THIS TREE
outfile
infile
outfile
infile
outfile
outfile
D:\WANCHENG\SOFTWARE\PHYLIP\PHYLIPWIN\CONSENSE.EXE: can't read infile
Please enter a new filename>D:\wancheng\software\Phylip\phylipwin\wan1
←[2J←[H
Majority-rule 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
0 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)
CONSENSUS TREE:
the numbers at the forks indicate the number
of times the group consisting of the species
which are to the right of that fork occurred
among the trees, out of 100.00 trees
THESE DISTANCE
remember: this is an unrooted tree!

CONSENSUS TREE:
the numbers on the branches indicate the number of times the partition of the species into the two sets which are separated by that branch occurred among the trees, out of 100.00 trees

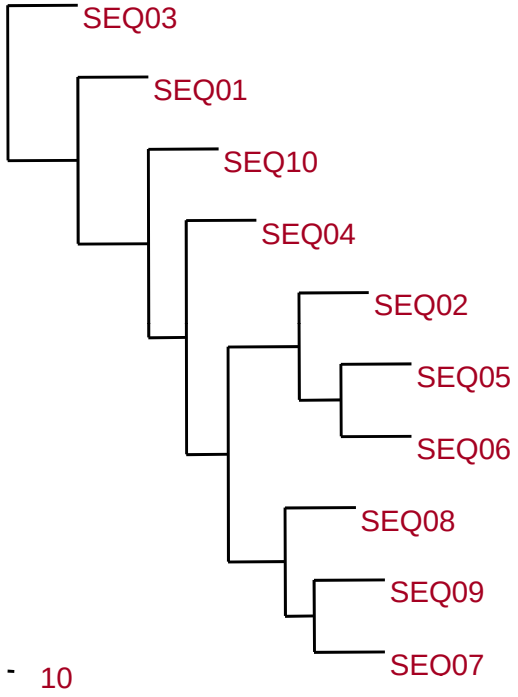




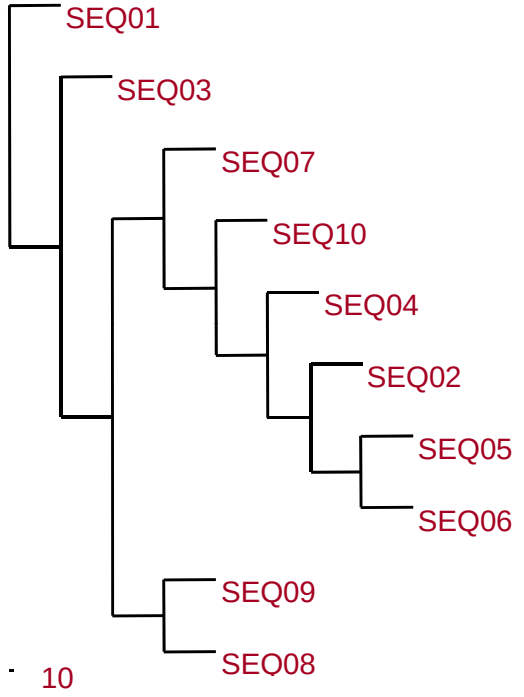
unrooted



VECTNTI Prediction



Distance Matrix Methods (NJ)



Character Matrix Methods (ML)

努力試 用力試 你就會了