Molecular Phylogenetic Analysis

PhylOgenetic WEb Repeater
Tips
Home
Demo
Help
Conta
PALM
Phylogenetic reconstruction by Automatic Likelihood Model selector
Home
Demo
Help
Contact

林仲彥

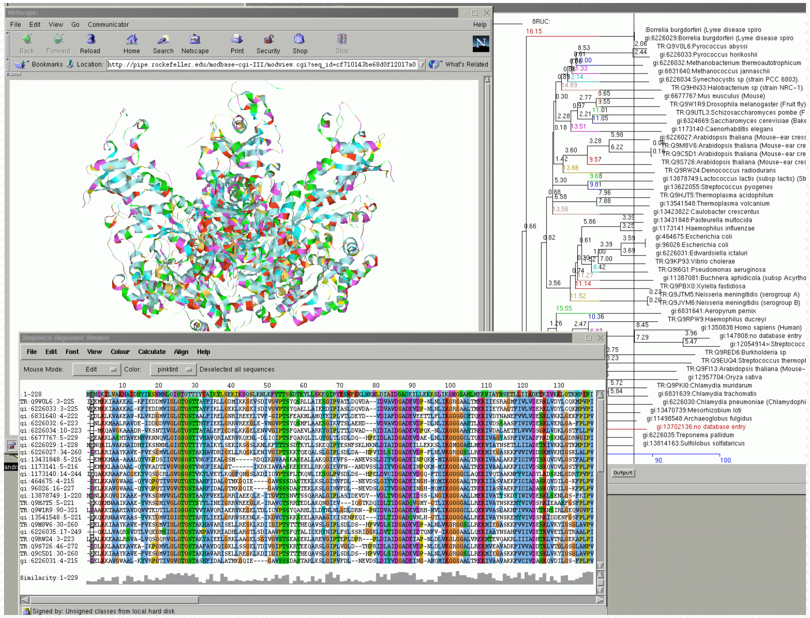
cylin@iis.sinica.edu.tw

LAB OF System Biology & Network Biology
中央研究院資訊科學研究所 @iis, Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室

Dec 4, 2009

http://eln.iis.sinica.edu.tw

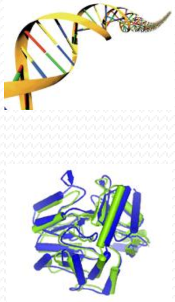
Coding Characters and Defining Homology


File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Shop Stop
Bookmarks Location: http://pipe.rockefeller.edu/modbase-cgi-III/modview.cgi?seq_id=cf710143be68d0f12017a0
What's Related
Sequence Alignment Window
File Edit Font View Colour Calculate Align Help
Mouse Mode: Edit Color: pinktint Deselected all sequences
Similarity:1-229
Signed by: Unsigned classes from local hard disk
Output

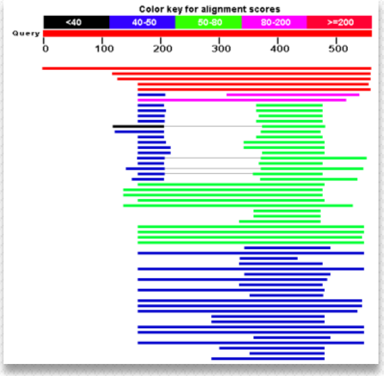
Classical phylogenetic analysis by Morphology

Molecular phylogenetic analysis By Bio-Molecules

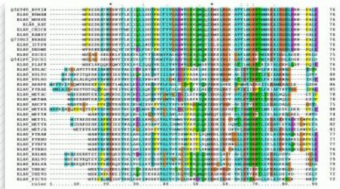
Steps of Phylogenetic Analysis



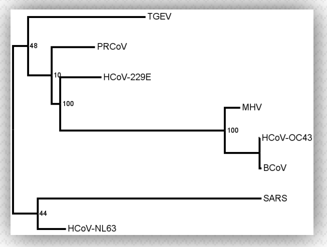
New Sequence



Homology Search



Alignment

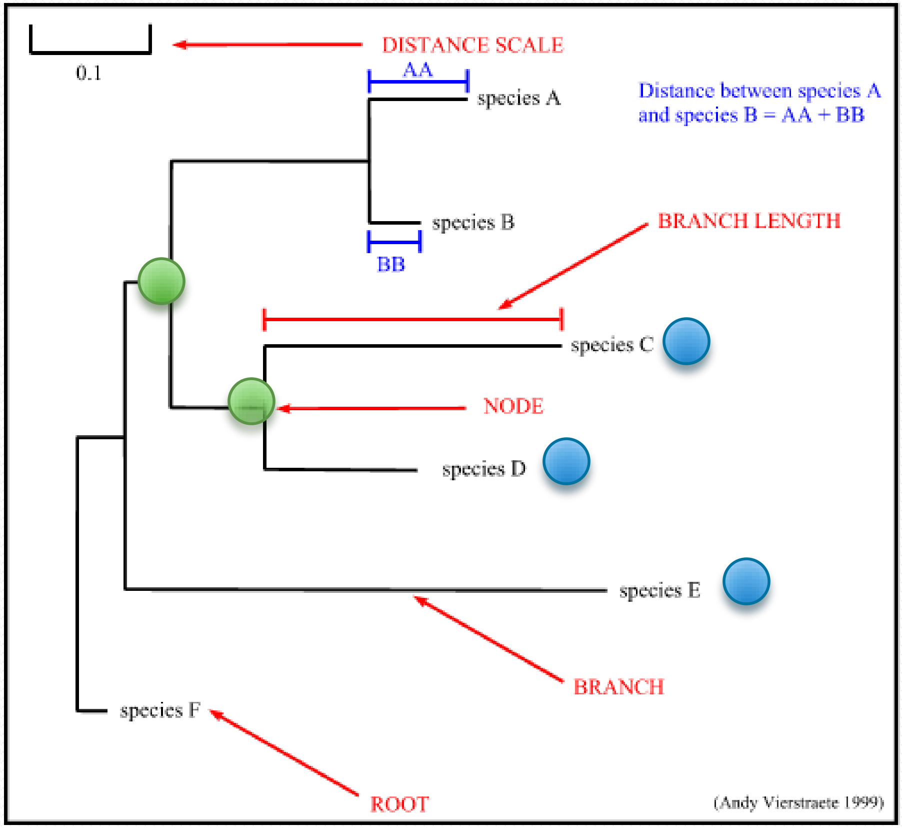


Phylogenetic Tree

Elements in a Phylogenetic Tree

- The tree is composed of nodes connected by branches.

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



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

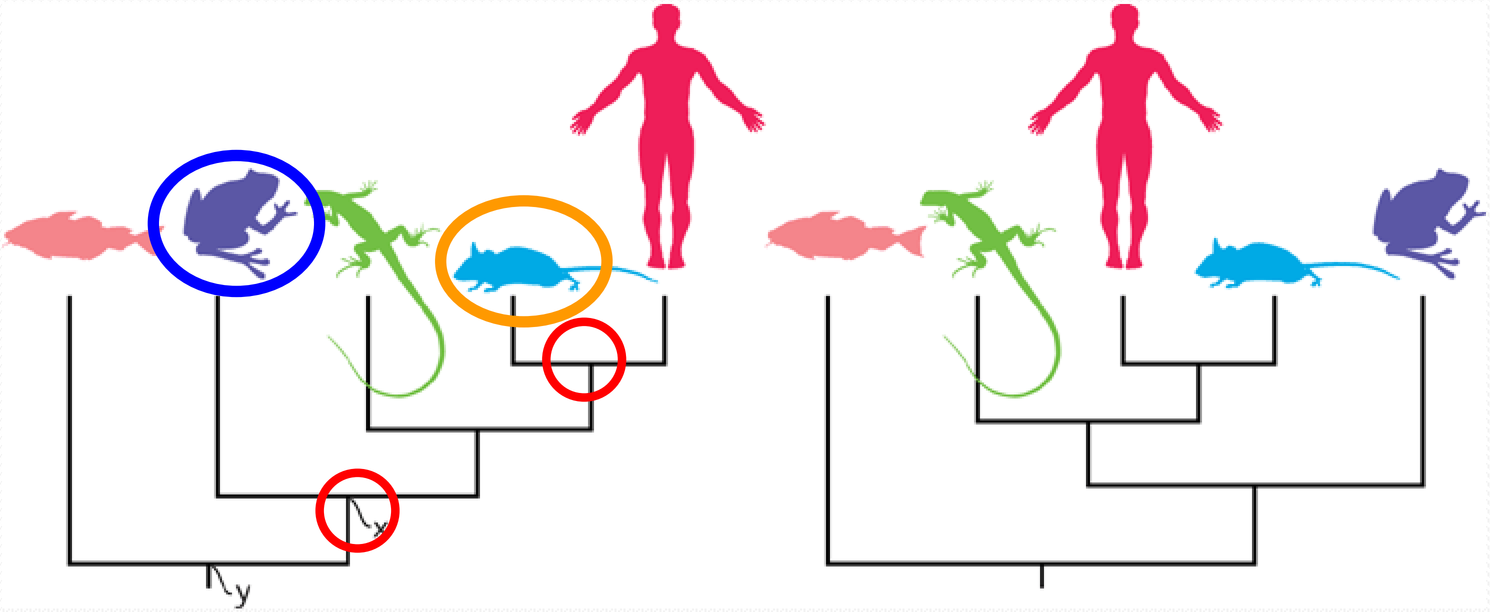
- **node :** a node represents a taxonomic unit.
  - Internal nodes 🔵
  - External nodes 🟢

- **root :** is the common ancestor of all taxa.

- **branch (edge):** defines the relationship between the taxa.

Trees Only Represent The Order Of Branching

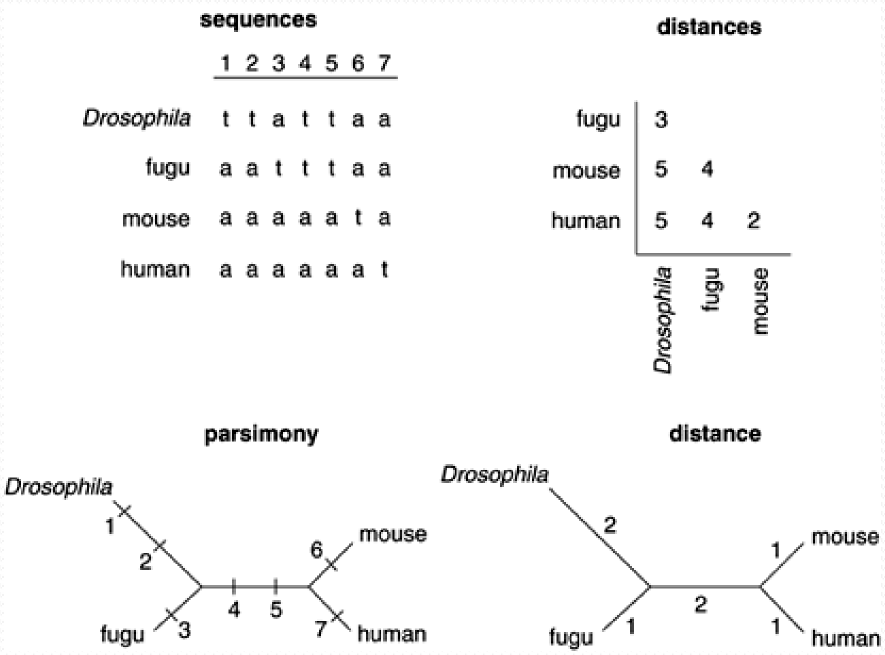
- Same topology in a different style
  - Both trees have identical topologies, with some of the internal nodes rotated.



(David A. Baum et al., Science 11 November 2005:Vol. 310. no. 5750, pp. 979 – 980)

The Ways to Construct the tree

- Distance-matrix methods
  - Neighbor-joining
  - Fitch-Margoliash method
  - Using outgroups
- Maximum parsimony
  - Branch and bound
  - MALIGN and POY
- Maximum likelihood (Statistics Based)
- Bayesian inference (Probability Based)



sequences

	1	2	3	4	5	6	7
Drosophila	t	t	a	t	t	a	a
fugu	a	a	t	t	t	a	a
mouse	a	a	a	a	a	t	a
human	a	a	a	a	a	a	t

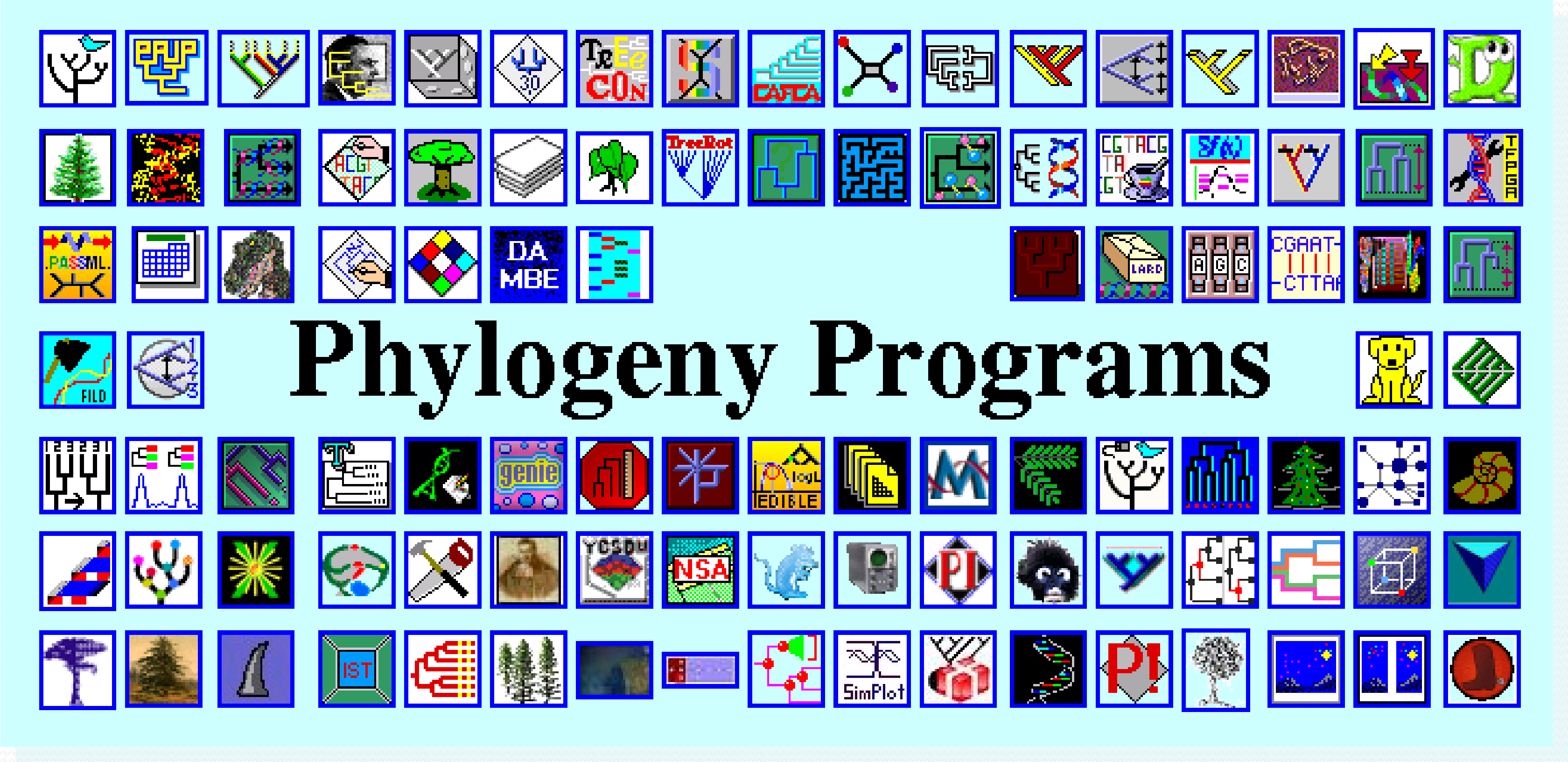
distances

	Drosophila	fugu	mouse
fugu	3		
mouse	5	4	
human	5	4	2

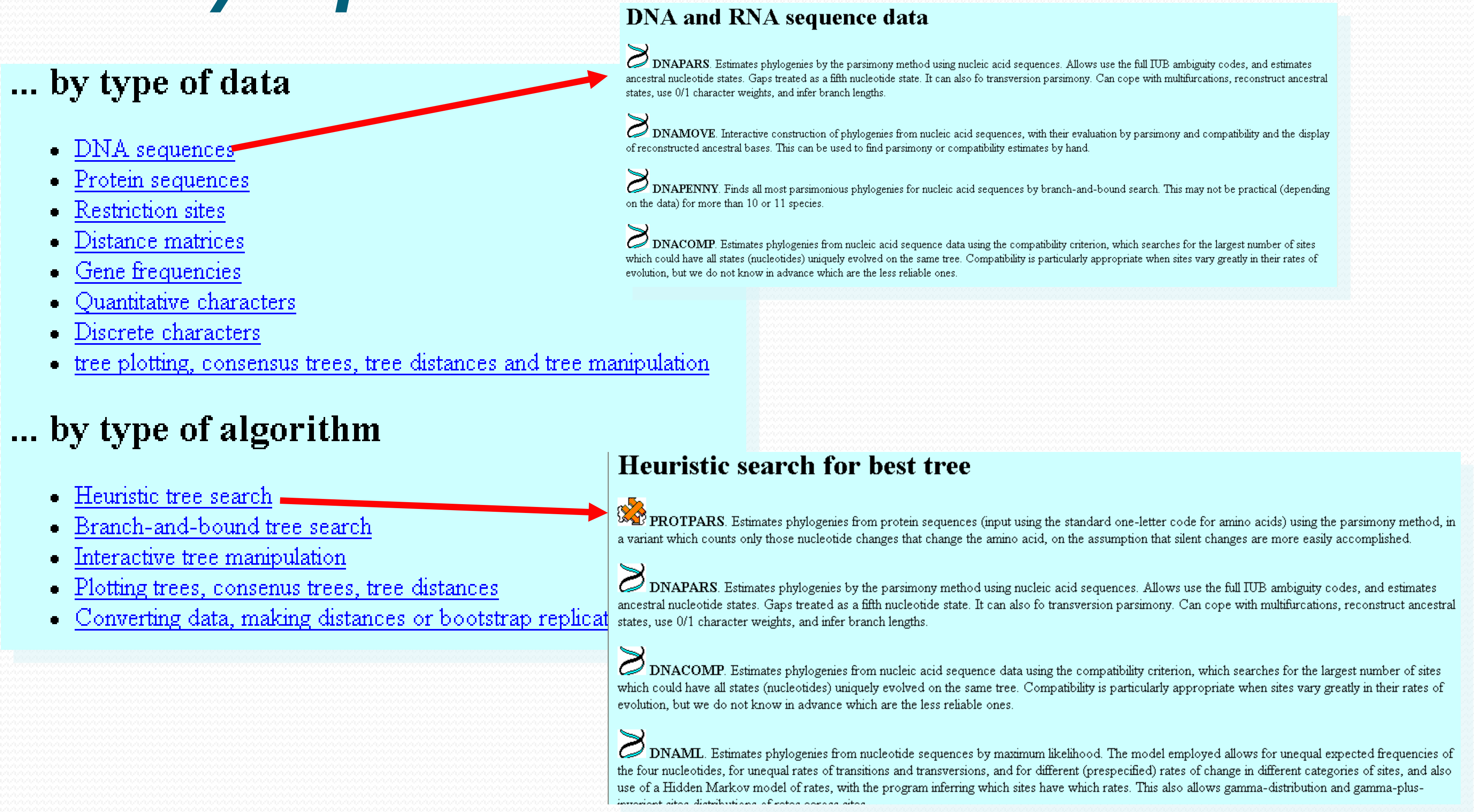
http://media.wiley.com/CurrentProtocols/BI/bi0601/bi0601-fig-0007-1-full.gif

Phylogeny Packages

http://evolution.genetics.washington.edu/phylip/software.html



Phylip



... 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 replicat

DNA and RNA sequence data

DNAPARS. Estimates phylogenies by the parsimony method using nucleic acid sequences. Allows use the full IUB ambiguity codes, and estimates ancestral nucleotide states. Gaps treated as a fifth nucleotide state. It can also fo transversion parsimony. Can cope with multifurcations, reconstruct ancestral states, use 0/1 character weights, and infer branch lengths.

DNAMOVE. Interactive construction of phylogenies from nucleic acid sequences, with their evaluation by parsimony and compatibility and the display of reconstructed ancestral bases. This can be used to find parsimony or compatibility estimates by hand.

DNAPENNY. Finds all most parsimonious phylogenies for nucleic acid sequences by branch-and-bound search. This may not be practical (depending on the data) for more than 10 or 11 species.

DNACOMP. Estimates phylogenies from nucleic acid sequence data using the compatibility criterion, which searches for the largest number of sites which could have all states (nucleotides) uniquely evolved on the same tree. Compatibility is particularly appropriate when sites vary greatly in their rates of evolution, but we do not know in advance which are the less reliable ones.

Heuristic search for best tree

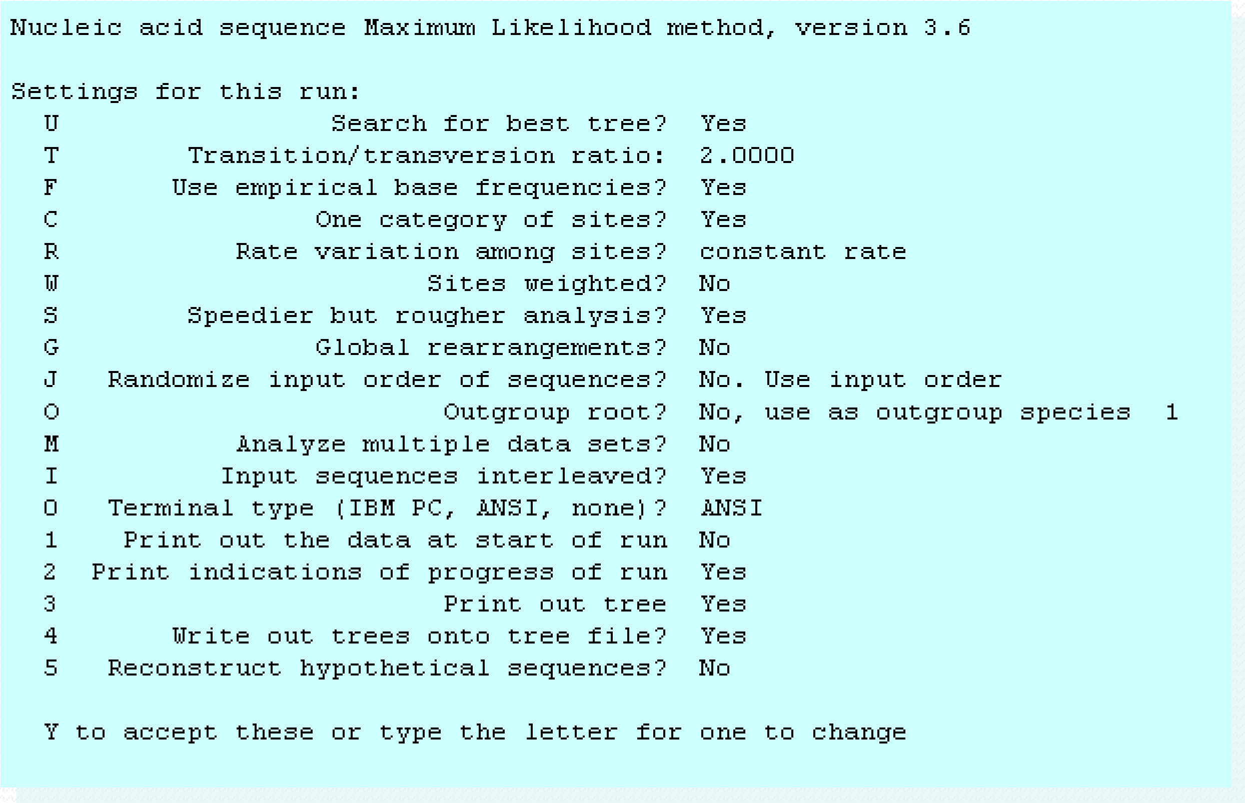
PROTPARS. Estimates phylogenies from protein sequences (input using the standard one-letter code for amino acids) using the parsimony method, in a variant which counts only those nucleotide changes that change the amino acid, on the assumption that silent changes are more easily accomplished.

DNAPARS. Estimates phylogenies by the parsimony method using nucleic acid sequences. Allows use the full IUB ambiguity codes, and estimates ancestral nucleotide states. Gaps treated as a fifth nucleotide state. It can also fo transversion parsimony. Can cope with multifurcations, reconstruct ancestral states, use 0/1 character weights, and infer branch lengths.

DNACOMP. Estimates phylogenies from nucleic acid sequence data using the compatibility criterion, which searches for the largest number of sites which could have all states (nucleotides) uniquely evolved on the same tree. Compatibility is particularly appropriate when sites vary greatly in their rates of evolution, but we do not know in advance which are the less reliable ones.

DNAML. Estimates phylogenies from nucleotide sequences by maximum likelihood. The model employed allows for unequal expected frequencies of the four nucleotides, for unequal rates of transitions and transversions, and for different (prespecified) rates of change in different categories of sites, and also use of a Hidden Markov model of rates, with the program inferring which sites have which rates. This also allows gamma-distribution and gamma-plus-invariant sites distributions of rates across sites.

Interactive Interface for Phylip

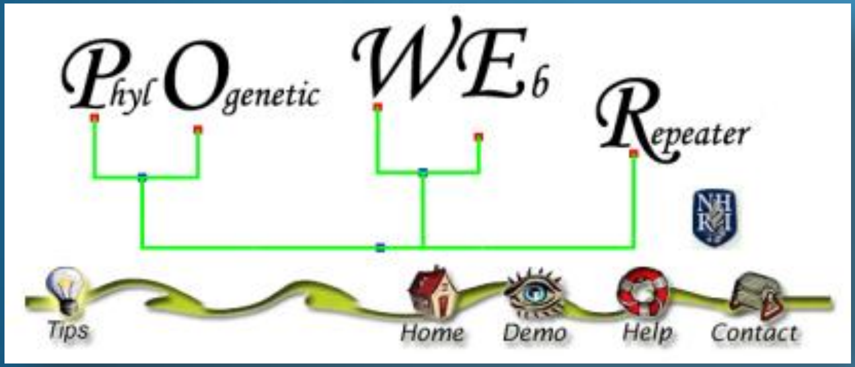


```
Nucleic acid sequence Maximum Likelihood method, version 3.6

Settings for this run:
  U                 Search for best tree?  Yes
  T        Transition/transversion ratio:  2.0000
  F       Use empirical base frequencies?  Yes
  C                One category of sites?  Yes
  R           Rate variation among sites?  constant rate
  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)?  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
  5   Reconstruct hypothetical sequences?  No

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

At this stage they do not have a mouse-windows interface for PHYLIP

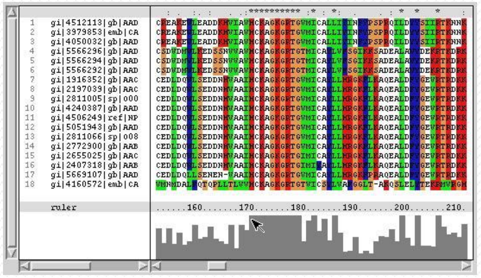
Phylogenetic Web Repeater (POWER):

An integrated and user-optimized framework for biomolecular phylogenetic analysis



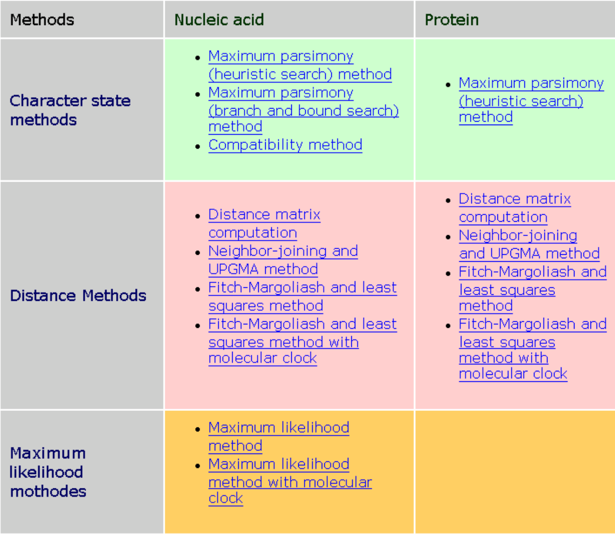
Nucleic Acids Research, 2005

General Pipeline for Phylogenetic Analysis



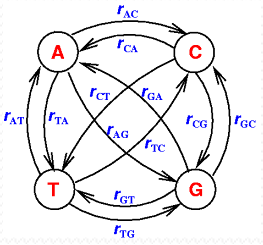
```
                         :  .  : . :   .  . :********** :* : *:  .     . : *  *   *   :
 1  gi|4512113|gb|AAD    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 2  gi|3979853|emb|CA    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 3  gi|4050032|gb|AAD    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 4  gi|5566296|gb|AAD    CSDVDMWLKEDSSNVVAVHCKAGKGRTGTMICAYLVFSGIKKSADEALAWYDEKRTKDRK
 5  gi|5566294|gb|AAD    CSDVDMWLKEDSSNVVAVHCKAGKGRTGTMICAYLVFSGIKKSADEALAWYDEKRTKDRK
 6  gi|5566292|gb|AAD    CSDVDMWLKEDSSNVVAVHCKAGKGRTGTMICAYLVFSGIKKSADEALAWYDEKRTKDRK
 7  gi|1916352|gb|AAC    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
 8  gi|2197039|gb|AAC    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
 9  gi|2811005|sp|O00    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
10  gi|4240387|gb|AAD    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
11  gi|4506249|ref|NP    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
12  gi|5051943|gb|AAD    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
13  gi|2811066|sp|O08    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
14  gi|2772900|gb|AAB    CEDLDQWLSEDDNHVAAIHCKAGKGRTGVMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
15  gi|2655025|gb|AAC    CEDLDQWLSEDDNHVAAIHCKAGKGRTGIMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
16  gi|2407318|gb|AAB    CEDLDQWLSEDDNHVAAIHCKAGKGRTGIMIYAYLLHRGKFLKAQEALDFYGEVRTRDKK
17  gi|5669107|gb|AAD    CEDLDQLLSENEN-VAAIHCKAGKGRTGVMICAYLLHRGKFPRAQEALDFYGEVRTRDKK
18  gi|4160572|emb|CA    VMNMDALFQTQPLLTLVVHCKAGKGRTGTVICSYLVAFGGLT-AKQSLELYTEKRMVRGH

    ruler                ......160.......170.......180.......190.......200.......210.
```

Multiple Sequence Alignment



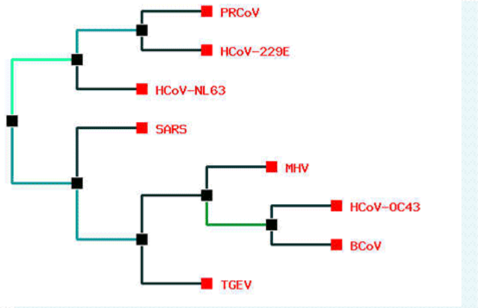
Methods	Nucleic acid	Protein
Character state methods	• Maximum parsimony (heuristic search) method • Maximum parsimony (branch and bound search) method • Compatibility method	• Maximum parsimony (heuristic search) method
Distance Methods	• Distance matrix computation • Neighbor-joining and UPGMA method • Fitch-Margoliash and least squares method • Fitch-Margoliash and least squares method with molecular clock	• Distance matrix computation • Neighbor-joining and UPGMA method • Fitch-Margoliash and least squares method • Fitch-Margoliash and least squares method with molecular clock
Maximum likelihood methodes	• Maximum likelihood method • Maximum likelihood method with molecular clock	

Selection of inference Methods



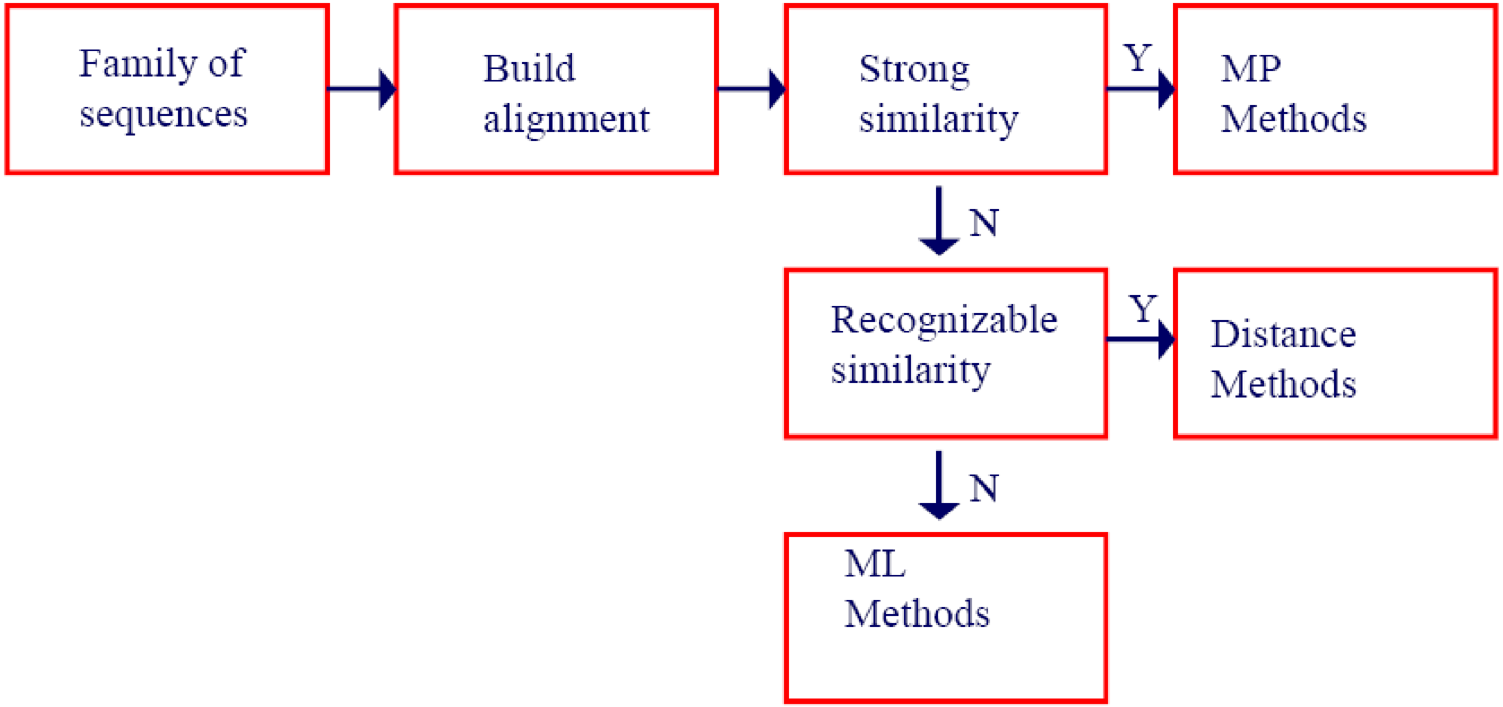
Substitution Model
Tree Construction

Bootstrapping



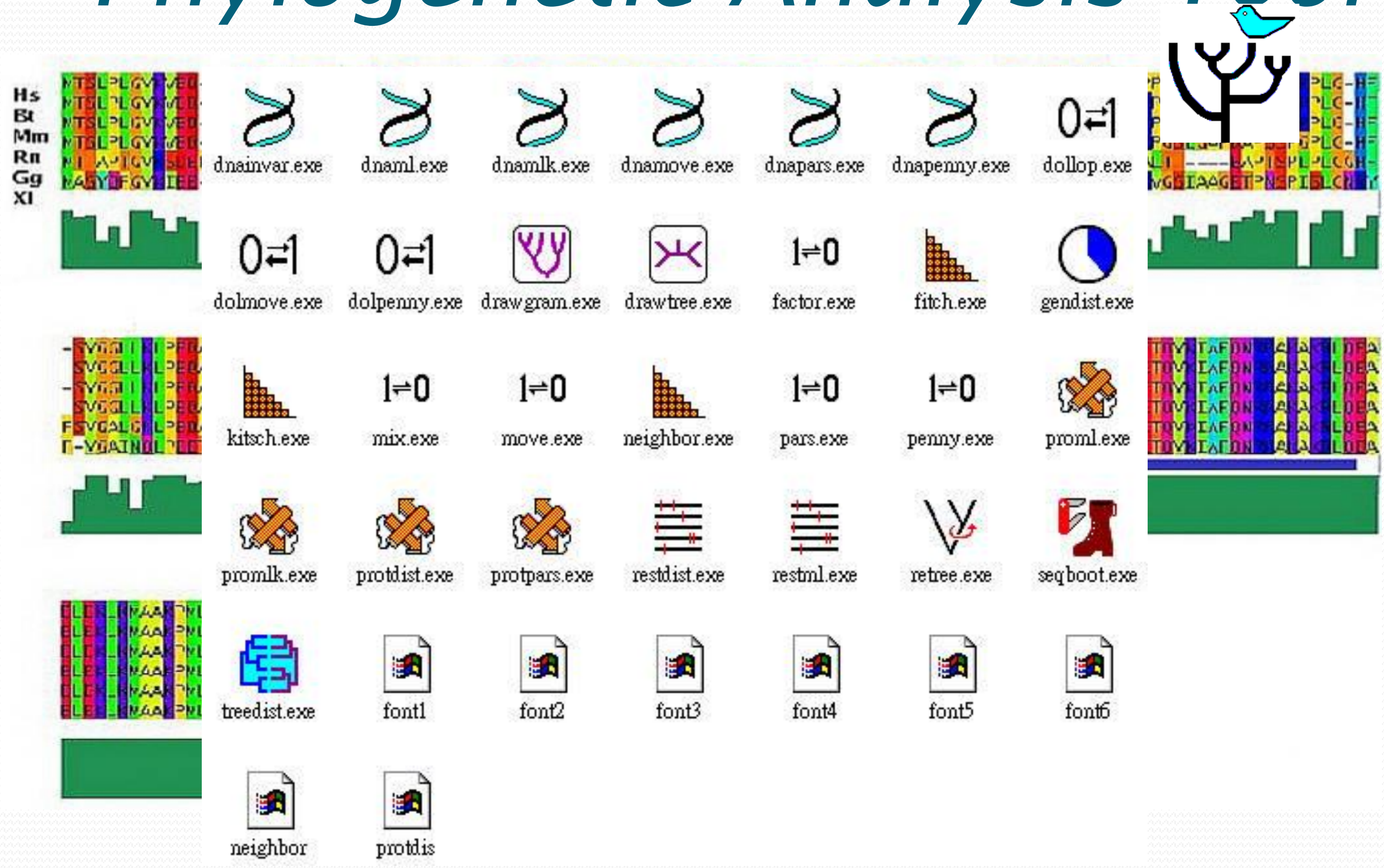
Evaluate phylogenetic tree

Flowchart of Analysis



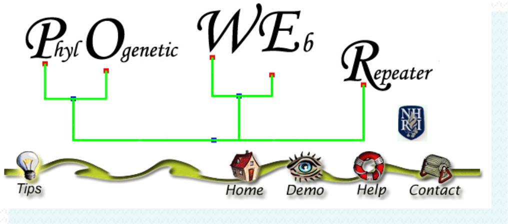
(Mount, *Bioinformatics*)

Phylogenetic Analysis Tool

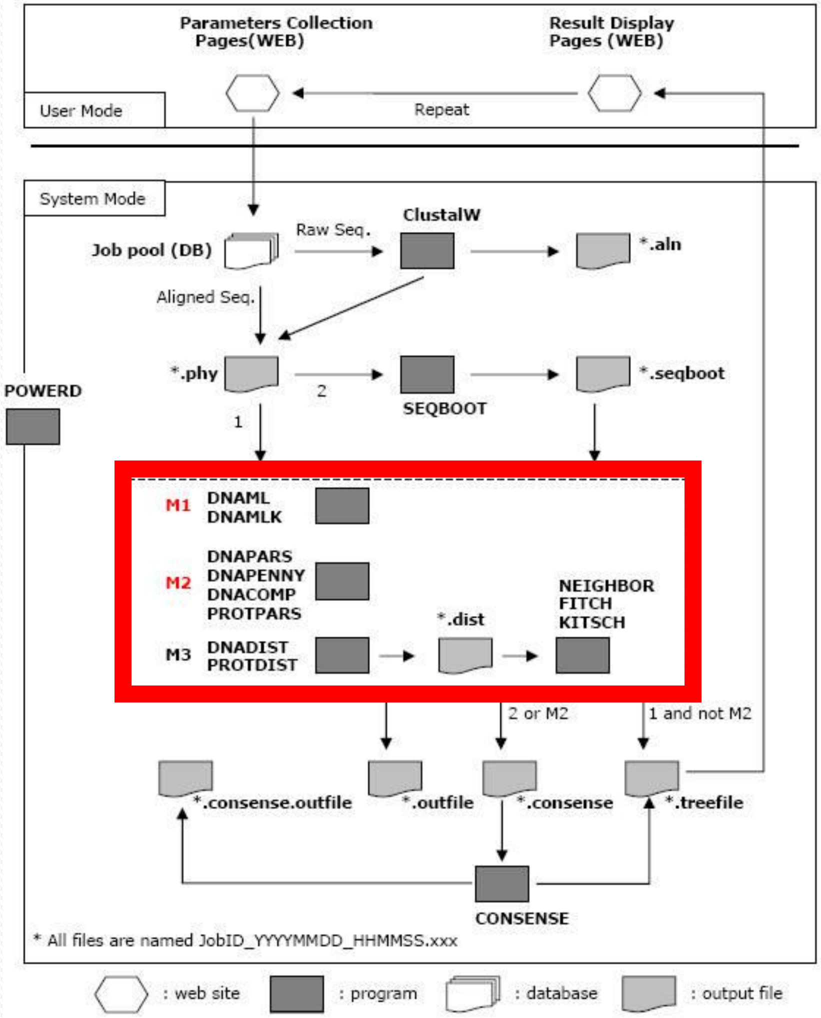


POWER: PhylOgenetic WEb Repeater

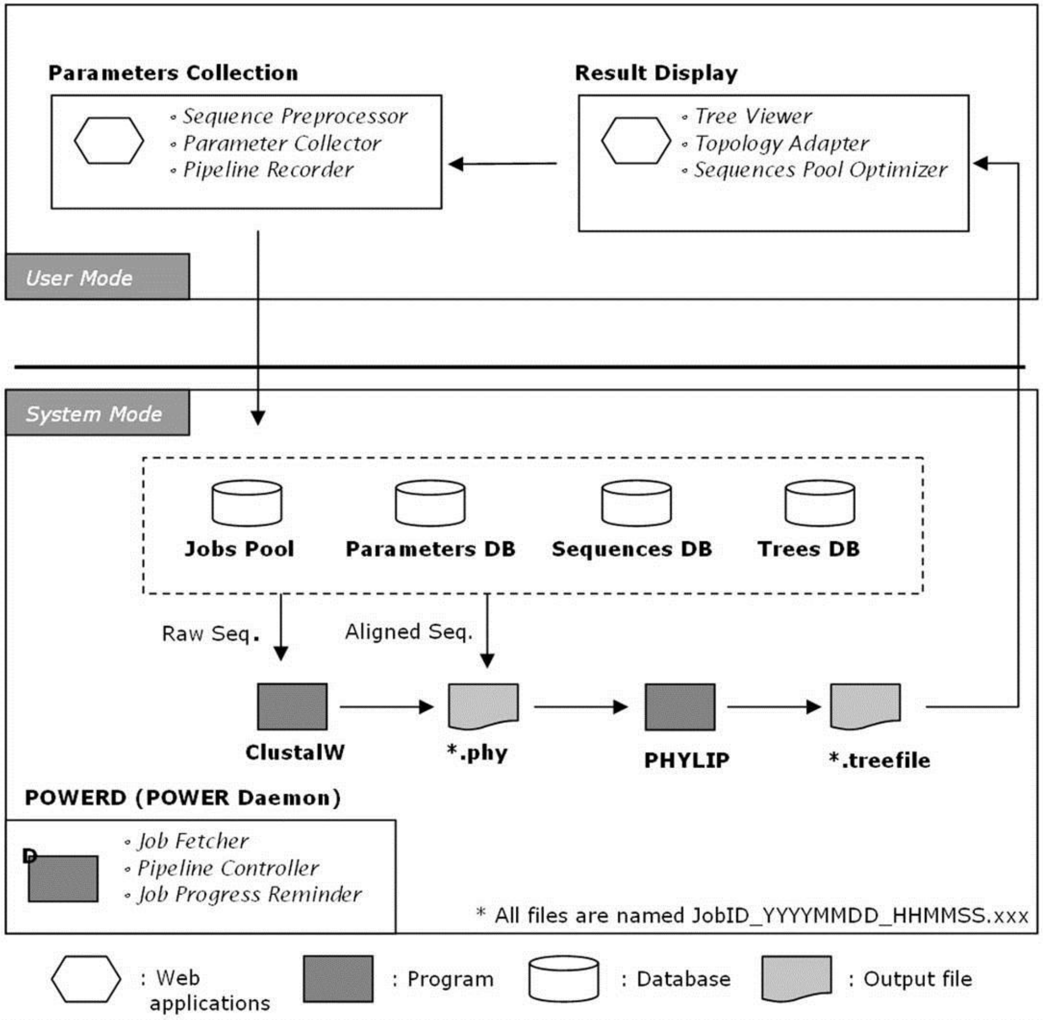
- Provide a seamless way to conduct the complex phylogenetic analysis for Biologists
- An integrated and user-optimized framework for biomolecular phylogenetic analysis
- POWER uses an open-source LAMP (Linux, Apache, MySQL, PHP) structure and infers genetic distances and phylogenetic relationships using well-established algorithms (ClustalW and PHYLIP)
- Through a user-friendly web interface, users can sketch a tree effortlessly in multiple steps
- Furthermore, iterative tree construction can be performed by adding sequences to, or removing them from, a previously submitted job



Integration of Phylip Packages into Automatic Flow



Inside of POWER



Parameters Collection

- Sequence Preprocessor
- Parameter Collector
- Pipeline Recorder

Result Display

- Tree Viewer
- Topology Adapter
- Sequences Pool Optimizer

User Mode

System Mode

Jobs Pool

Parameters DB

Sequences DB

Trees DB

Raw Seq.

Aligned Seq.

ClustalW

*.phy

PHYLIP

*.treefile

POWERD (POWER Daemon)

D

- Job Fetcher
- Pipeline Controller
- Job Progress Reminder

* All files are named JobID_YYYYMMDD_HHMMSS.xxx

: Web applications

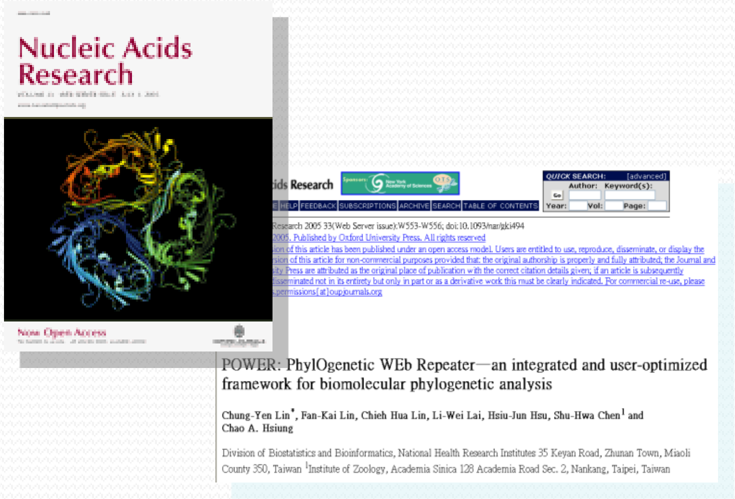
: Program

: Database

: Output file

POWER: PhylOgenetic WEb Repeater

http://power.nhri.org.tw



Research 2005 33(Web Server issue):W553-W556; doi:10.1093/nar/gki494

2005. Published by Oxford University Press. All rights reserved

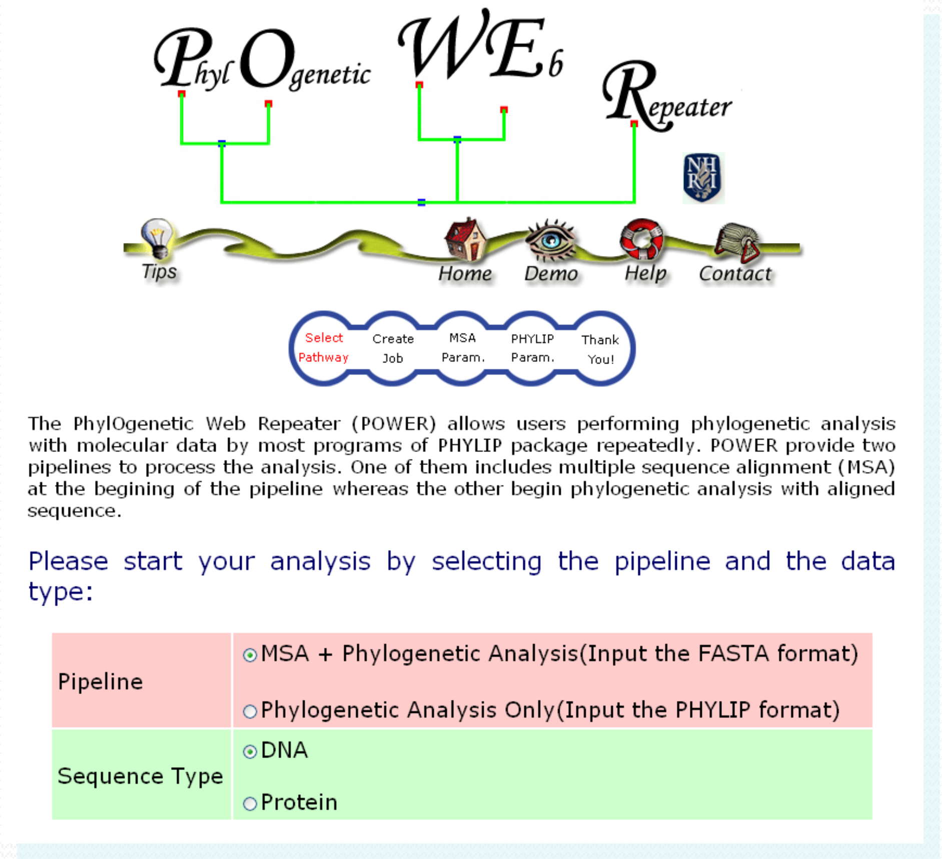
sion of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the
sion of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and
sity Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently
disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use, please
.permissions[at]oupjournals.org

POWER: PhylOgenetic WEb Repeater—an integrated and user-optimized framework for biomolecular phylogenetic analysis

Chung-Yen Lin*, Fan-Kai Lin, Chieh Hua Lin, Li-Wei Lai, Hsiu-Jun Hsu, Shu-Hwa Chen[1] and Chao A. Hsiung

Division of Biostatistics and Bioinformatics, National Health Research Institutes 35 Keyan Road, Zhunan Town, Miaoli County 350, Taiwan [1]Institute of Zoology, Academia Sinica 128 Academia Road Sec. 2, Nankang, Taipei, Taiwan

Nucl. Acids Res. 2005 33: W553-W556



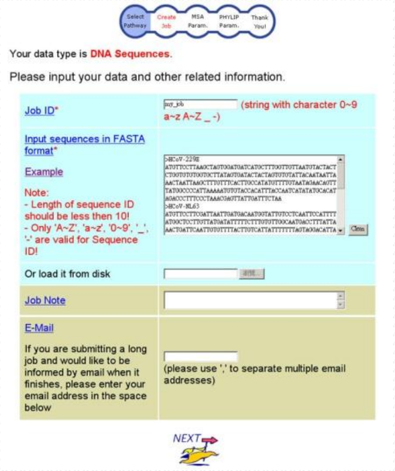
The PhylOgenetic Web Repeater (POWER) allows users performing phylogenetic analysis with molecular data by most programs of PHYLIP package repeatedly. POWER provide two pipelines to process the analysis. One of them includes multiple sequence alignment (MSA) at the begining of the pipeline whereas the other begin phylogenetic analysis with aligned sequence.

Please start your analysis by selecting the pipeline and the data type:

Pipeline	● MSA + Phylogenetic Analysis(Input the FASTA format) ○ Phylogenetic Analysis Only(Input the PHYLIP format)
Sequence Type	● DNA ○ Protein

PhylOgenetic Web Repeater (POWER)

Data Input

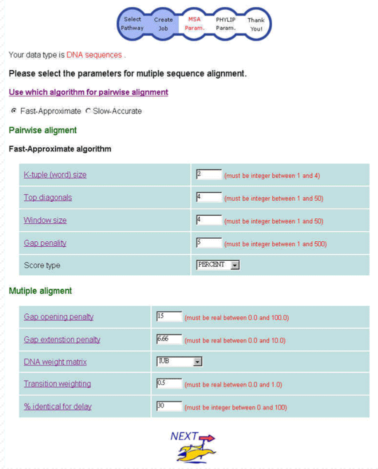


Your data type is **DNA Sequences.**

Please input your data and other related information.

Job ID*	my_job (string with character 0~9 a~z A~Z _ -)
Input sequences in FASTA format* Example Note: - Length of sequence ID should be less then 10! - Only 'A~Z', 'a~z', '0~9', '_', '-' are valid for Sequence ID!	>HCoV-229E ATGTTCCTTAAGCTAGTGGATGATCATGCTTTGGTTGTTAATGTACTACT CTGGTGTGTGGTGCTTATAGTGATACTACTAGTGTGTATTACAATAATTA AACTAATTAAGCTTTGTTTCACTTGCCATATGTTTTGTAATAGAACAGTT TATGGCCCCATTAAAAATGTGTACCACATTTACCAATCATATATGCACAT AGACCCTTTCCCTAAACGAGTTATTGATTTCTAA >HCoV-NL63 ATGTTCCTTCGATTAATTGATGACAATGGTATTGTCCTCAATTCCATTTT ATGGCTCCTTGTTATGATATTTTTCTTTGTGTTGGCAATGACCTTTATTA AACTGATTCAATTGTGTTTTACTTGTCATTATTTTTTTAGTAGGACATTA Clean
Or load it from disk	
Job Note	
E-Mail If you are submitting a long job and would like to be informed by email when it finishes, please enter your email address in the space below	(please use ',' to separate multiple email addresses)

MSA parameter selection



Your data type is DNA sequences .

Please select the parameters for mutiple sequence alignment.

Use which algorithm for pairwise alignment

● Fast-Approximate ○ Slow-Accurate

Pairwise aligment

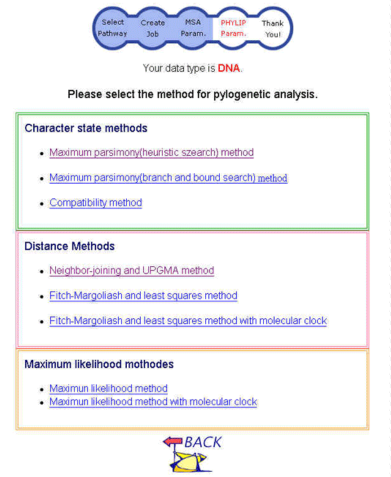
Fast-Approximate algorithm

K-tuple (word) size	2 (must be integer between 1 and 4)
Top diagonals	4 (must be integer between 1 and 50)
Window size	4 (must be integer between 1 and 50)
Gap penality	5 (must be integer between 1 and 500)
Score type	PERCENT

Mutiple aligment

Gap opening penalty	15 (must be real between 0.0 and 100.0)
Gap extenstion penalty	6.66 (must be real between 0.0 and 10.0)
DNA weight matrix	IUB
Transition weighting	0.5 (must be real between 0.0 and 1.0)
% identical for delay	30 (must be integer between 0 and 100)

Phylogeny inference



Your data type is **DNA**.

Please select the method for pylogenetic analysis.

Character state methods

- Maximum parsimony(heuristic szearch) method
- Maximum parsimony(branch and bound search) method
- Compatibility method

Distance Methods

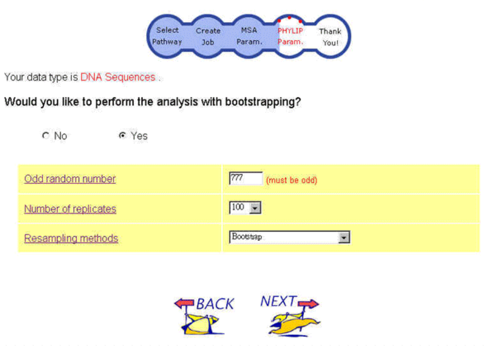
- Neighbor-joining and UPGMA method
- Fitch-Margoliash and least squares method
- Fitch-Margoliash and least squares method with molecular clock

Maximum likelihood mothodes

- Maximun likelihood method
- Maximun likelihood method with molecular clock

PhylOgenetic Web Repeater (POWER)

Options of bootstrapping



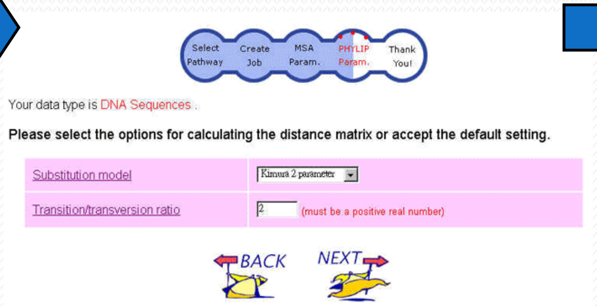
Your data type is DNA Sequences .

Would you like to perform the analysis with bootstrapping?

○ No ● Yes

Odd random number	777 (must be odd)
Number of replicates	100
Resampling methods	Bootstrap

Selection of substitution model

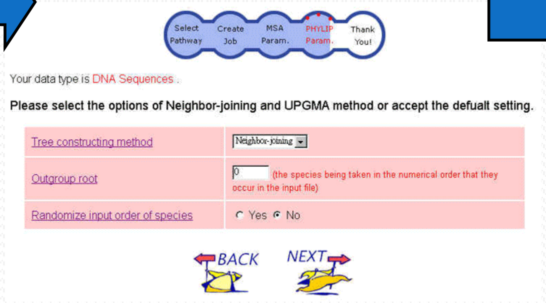


Your data type is DNA Sequences .

Please select the options for calculating the distance matrix or accept the default setting.

Substitution model	Kimura 2 parameter
Transition/transversion ratio	2 (must be a positive real number)

Selected method for phylogeny inference



Your data type is DNA Sequences .

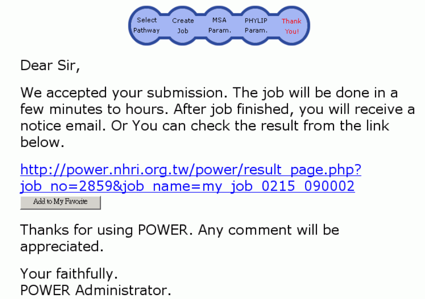
Please select the options of Neighbor-joining and UPGMA method or accept the defualt setting.

Tree constructing method	Neighbor-joining
Outgroup root	0 (the species being taken in the numerical order that they occur in the input file)
Randomize input order of species	○ Yes ● No

PhylOgenetic Web Repeater (POWER)

Result and Logs

Online or as bookmark



Dear Sir,

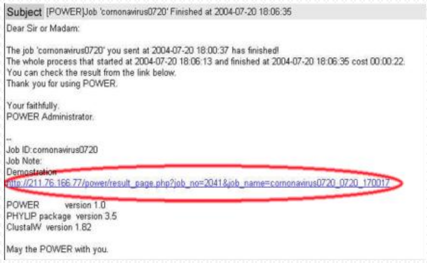
We accepted your submission. The job will be done in a few minutes to hours. After job finished, you will receive a notice email. Or You can check the result from the link below.

http://power.nhri.org.tw/power/result_page.php?job_no=2859&job_name=my_job_0215_090002

Thanks for using POWER. Any comment will be appreciated.

Your faithfully.
POWER Administrator.

Or E-mail notification



Subject [POWER]Job 'cornonavirus0720' Finished at 2004-07-20 18:06:35

Dear Sir or Madam:

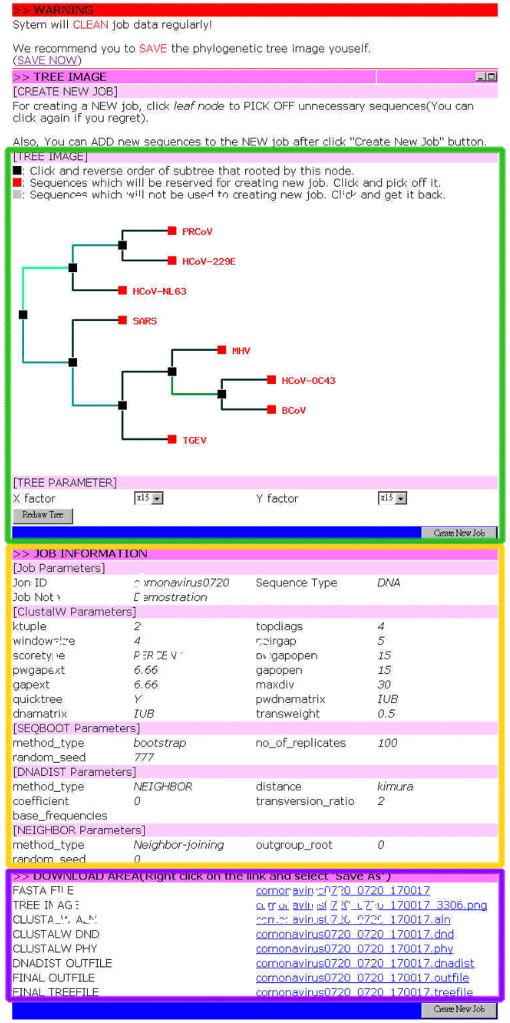
The job 'cornonavirus0720' you sent at 2004-07-20 18:00:37 has finished!
The whole process that started at 2004-07-20 18:06:13 and finished at 2004-07-20 18:06:35 cost 00:00:22.
You can check the result from the link below.
Thank you for using POWER.

Your faithfully.
POWER Administrator.

--
Job ID:cornonavirus0720
Job Note:
Demostration
http://211.76.166.77/power/result_page.php?job_no=2041&job_name=cornonavirus0720_0720_170017

POWER version 1.0
PHYLIP package version 3.5
ClustalW version 1.82

May the POWER with you.



>> WARNING

Sytem will CLEAN job data regularly!

We recommend you to SAVE the phylogenetic tree image youself.
(SAVE NOW)

>> TREE IMAGE

[CREATE NEW JOB]

For creating a NEW job, click *leaf node* to PICK OFF unnecessary sequences(You can click again if you regret).

Also, You can ADD new sequences to the NEW job after click "Create New Job" button.

[TREE IMAGE]

■: Click and reverse order of subtree that rooted by this node.
🟥: Sequences which will be reserved for creating new job. Click and pick off it.
▨: Sequences which will not be used to creating new job. Click and get it back.

[TREE PARAMETER]

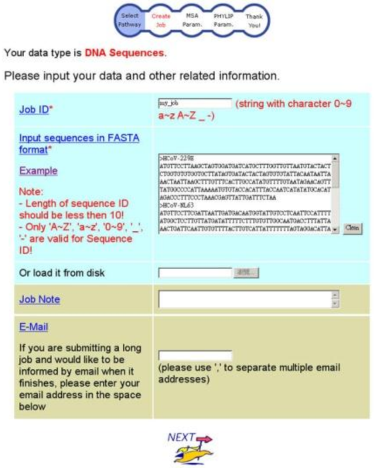
X factor	x15	Y factor	x15

>> JOB INFORMATION

[Job Parameters]			
Job ID	*cornonavirus0720*	Sequence Type	*DNA*
Job Note	*Demostration*		
[ClustalW Parameters]			
ktuple	*2*	topdiags	*4*
windowsize	*4*	pairgap	*5*
scoretype	*PERCENT*	pwgapopen	*15*
pwgapext	*6.66*	gapopen	*15*
gapext	*6.66*	maxdiv	*30*
quicktree	*Y*	pwdnamatrix	*IUB*
dnamatrix	*IUB*	transweight	*0.5*
[SEQBOOT Parameters]			
method_type	*bootstrap*	no_of_replicates	*100*
random_seed	*777*		
[DNADIST Parameters]			
method_type	*NEIGHBOR*	distance	*kimura*
coefficient	*0*	transversion_ratio	*2*
base_frequencies			
[NEIGHBOR Parameters]			
method_type	*Neighbor-joining*	outgroup_root	*0*
random_seed	*0*		

>> DOWNLOAD AREA(Right click on the link and select 'Save As')

FASTA FILE	cornonavirus0720_0720_170017
TREE IMAGE	cornonavirus0720_0720_170017_3306.png
CLUSTALW ALN	cornonavirus0720_0720_170017.aln
CLUSTALW DND	cornonavirus0720_0720_170017.dnd
CLUSTALW PHY	cornonavirus0720_0720_170017.phy
DNADIST OUTFILE	cornonavirus0720_0720_170017.dnadist
FINAL OUTFILE	cornonavirus0720_0720_170017.outfile
FINAL TREEFILE	cornonavirus0720_0720_170017.treefile



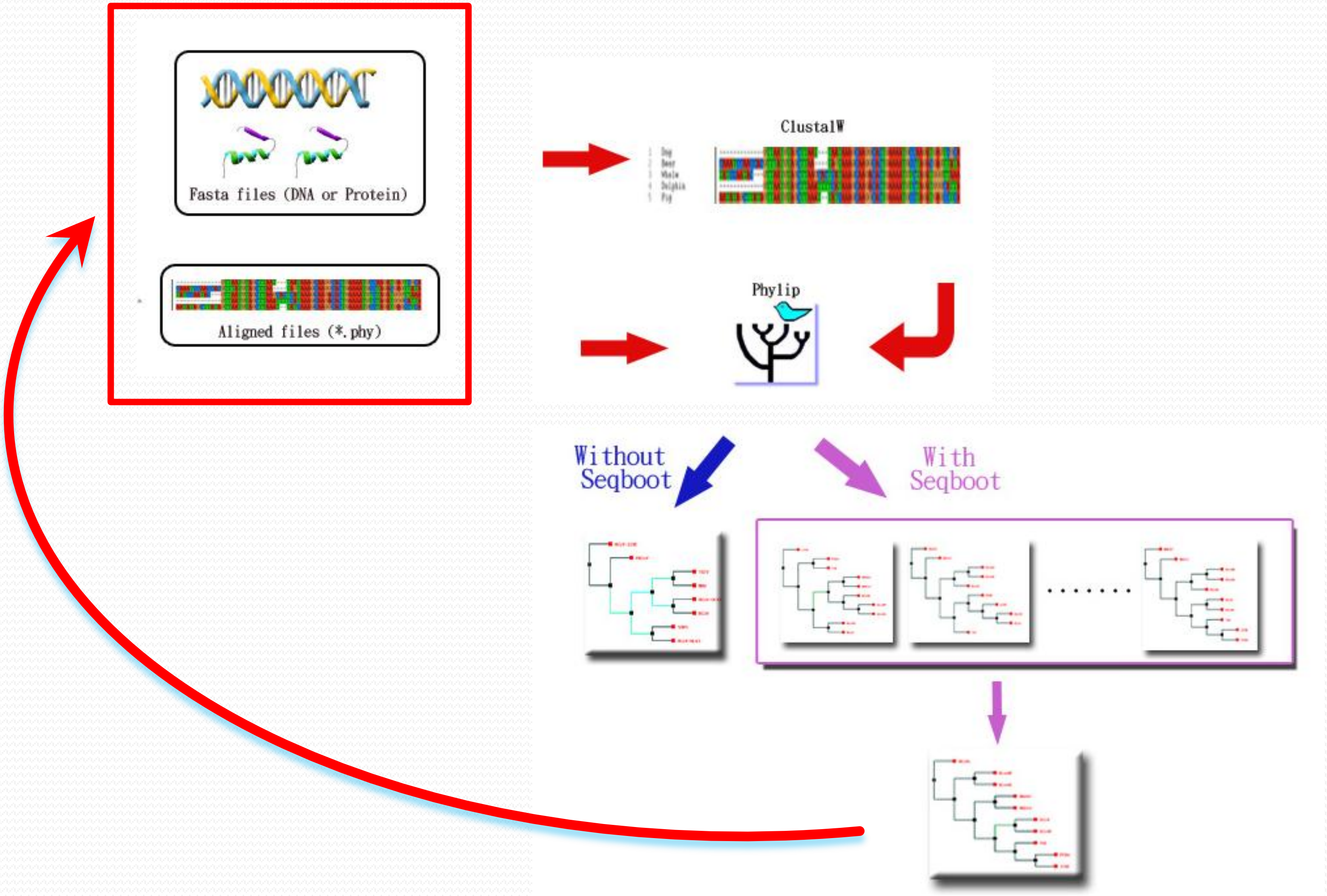
Your data type is **DNA Sequences.**

Please input your data and other related information.

Job ID*	my_job (string with character 0~9 a~z A~Z _ -)
Input sequences in FASTA format* Example Note: - Length of sequence ID should be less then 10! - Only 'A~Z', 'a~z', '0~9', '_', '-' are valid for Sequence ID!	>HCoV-229E ATGTTCCTTAAGCTAGTGGATGATCATGCTTTGGTTGTTAATGTACTACT CTGGTGTGTGGTGCTTATAGTGATACTACTAGTGTGTATTACAATAATTA AACTAATTAAGCTTTGTTTCACTTGCCATATGTTTTGTAATAGAACAGTT TATGGCCCCATTAAAAATGTGTACCACATTTACCAATCATATATGCACAT AGACCCTTTCCCTAAACGAGTTATTGATTTCTAA >HCoV-NL63 ATGTTCCTTCGATTAATTGATGACAATGGTATTGTCCTCAATTCCATTTT ATGGCTCCTTGTTATGATATTTTTCTTTGTGTTGGCAATGACCTTTATTA AACTGATTCAATTGTGTTTTACTTGTCATTATTTTTTTAGTAGGACATTA
Or load it from disk	
Job Note	
E-Mail If you are submitting a long job and would like to be informed by email when it finishes, please enter your email address in the space below	(please use ',' to separate multiple email addresses)

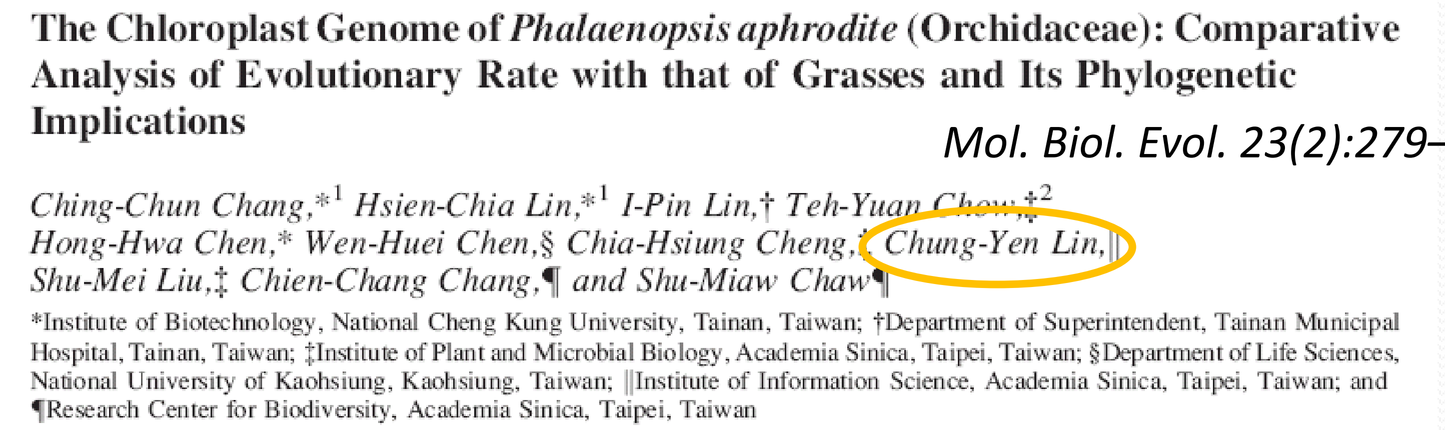
Re-perform the process by items added or deleted

PhylOgenetic Web Repeater (POWER)



Add/ delete sequences to invoke new job

Publication in POWER

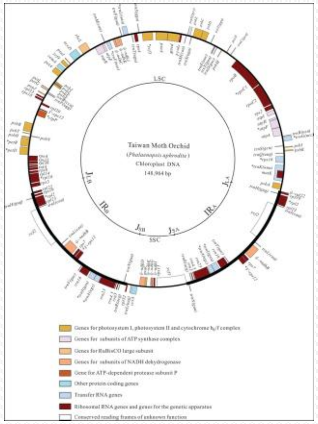


The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with that of Grasses and Its Phylogenetic Implications

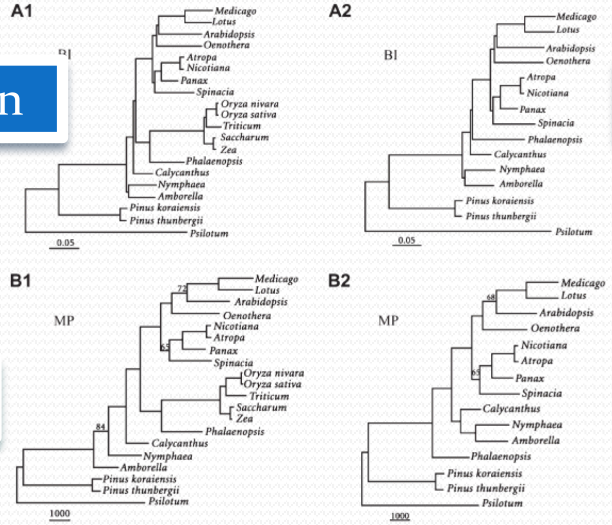
Mol. Biol. Evol. 23(2):279–291. 2006

Ching-Chun Chang,[1] Hsien-Chia Lin,*[1] I-Pin Lin,† Teh-Yuan Chow,‡[2] Hong-Hwa Chen,* Wen-Huei Chen,§ Chia-Hsiung Cheng,‡ Chung-Yen Lin,‖ Shu-Mei Liu,‡ Chien-Chang Chang,¶ and Shu-Miaw Chaw¶*

*Institute of Biotechnology, National Cheng Kung University, Tainan, Taiwan; †Department of Superintendent, Tainan Municipal Hospital, Tainan, Taiwan; ‡Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan; §Department of Life Sciences, National University of Kaohsiung, Kaohsiung, Taiwan; ‖Institute of Information Science, Academia Sinica, Taipei, Taiwan; and ¶Research Center for Biodiversity, Academia Sinica, Taipei, Taiwan


Taiwan Moth Orchid
(Phalaenopsis aphrodite)
Chloroplast DNA
148,964 bp
LSC
SSC
IRA
IRB
JLA
JLB
JSA
JSB
Genes for photosystem I, photosystem II and cytochrome b6/f complex
Genes for subunits of ATP synthase complex
Genes for RuBisCO large subunit
Genes for subunits of NADH dehydrogenase
Gene for ATP-dependent protease subunit P
Other protein coding genes
Transfer RNA genes
Ribosomal RNA genes and genes for the genetic apparatus
Conserved reading frames of unknown function

Bayesian



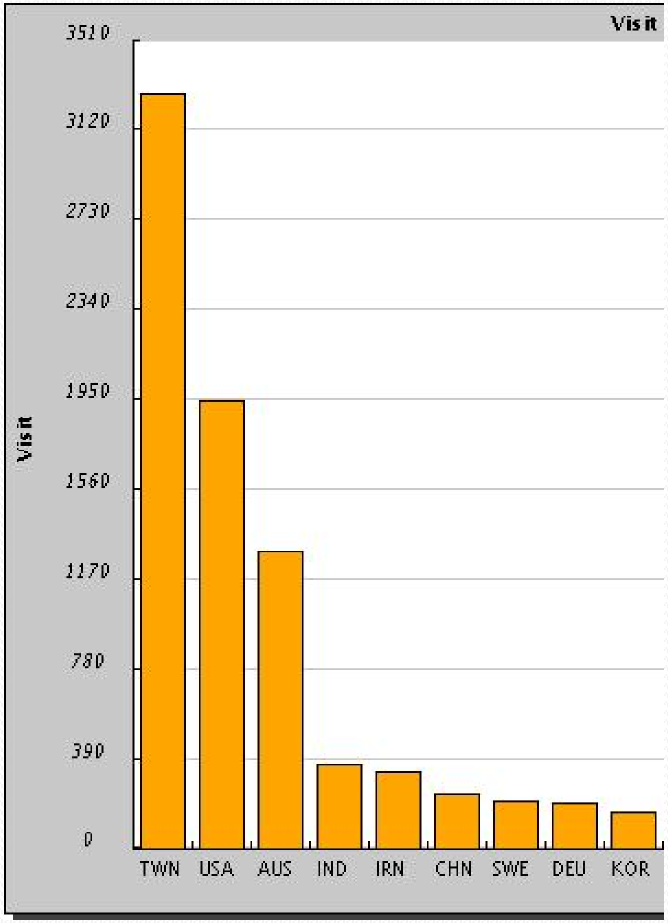
Bayesian

MP

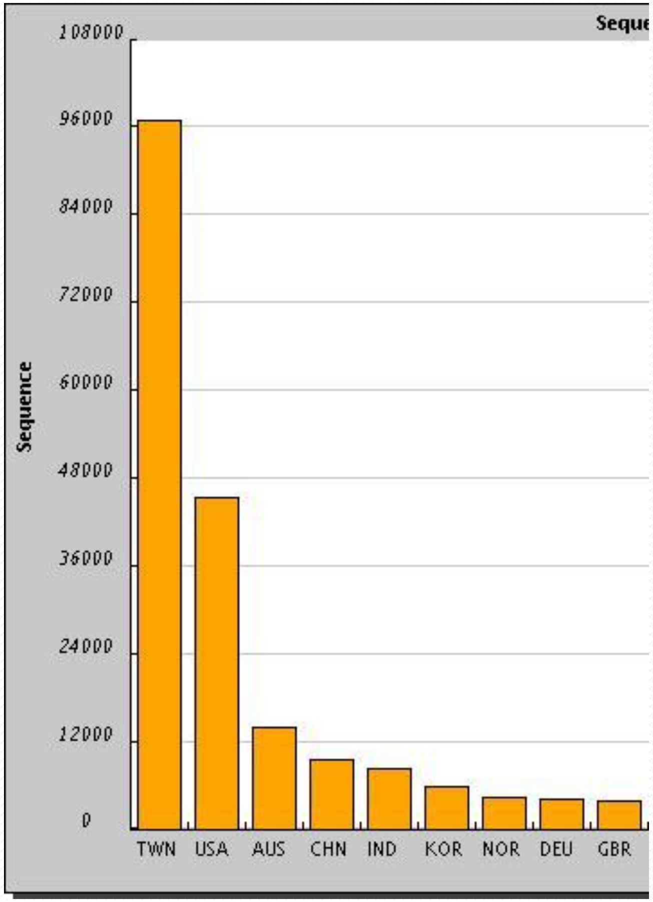
MP

Service Usage of POWER *from 2005 July.*

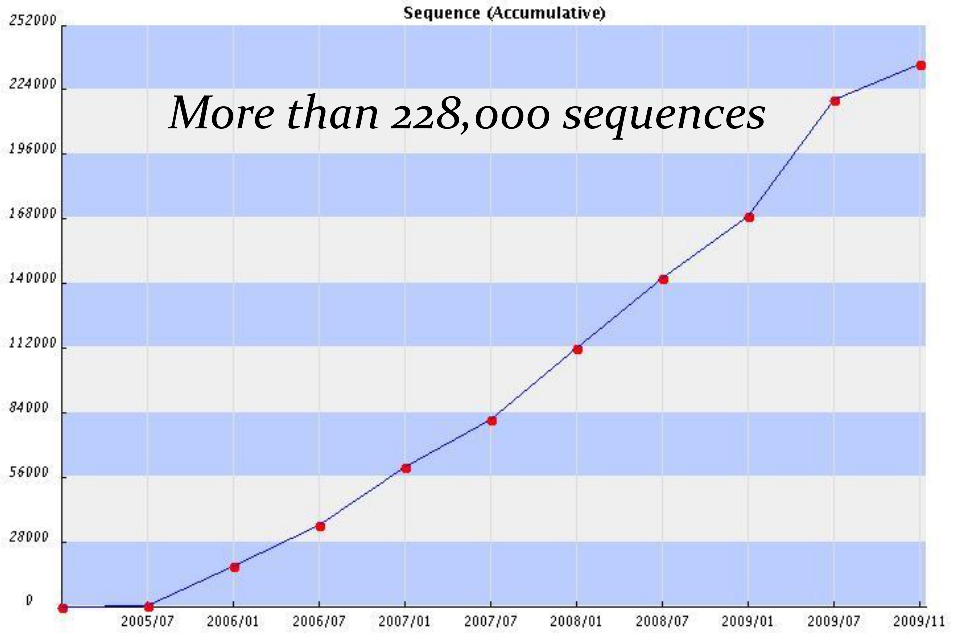
Accumulative Visit by Country



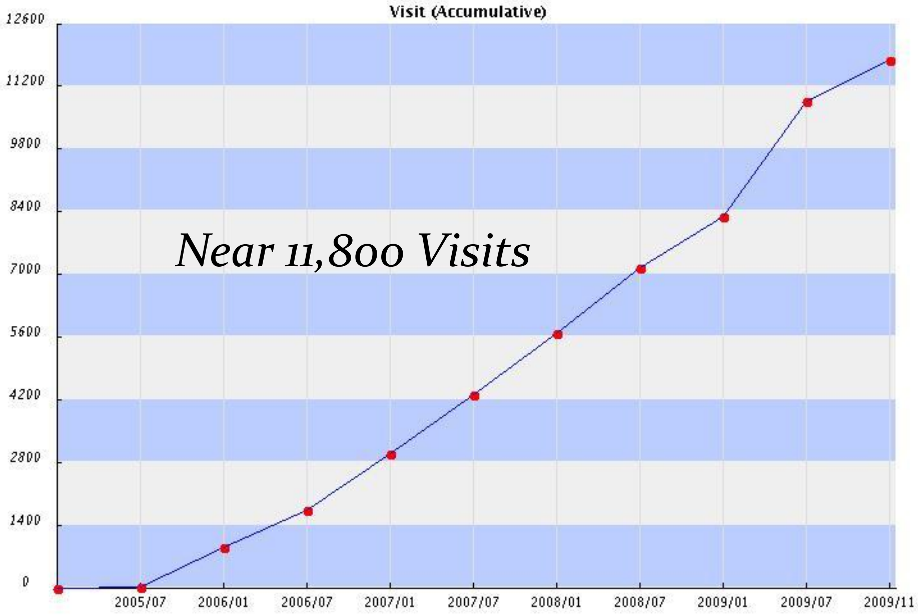
Accumulative Sequences by Country



Service Usage of POWER from 2005 July.

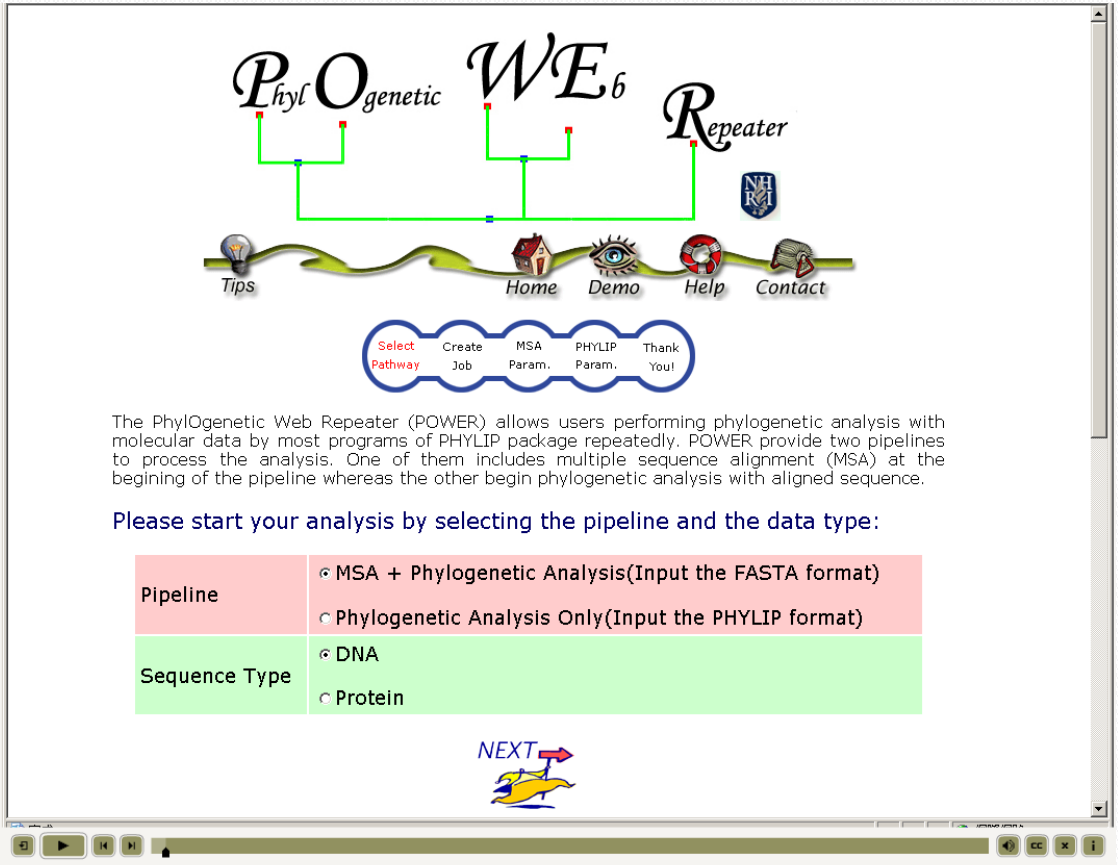


More than 228,000 sequences



Near 11,800 Visits

Automatic Online Demonstration



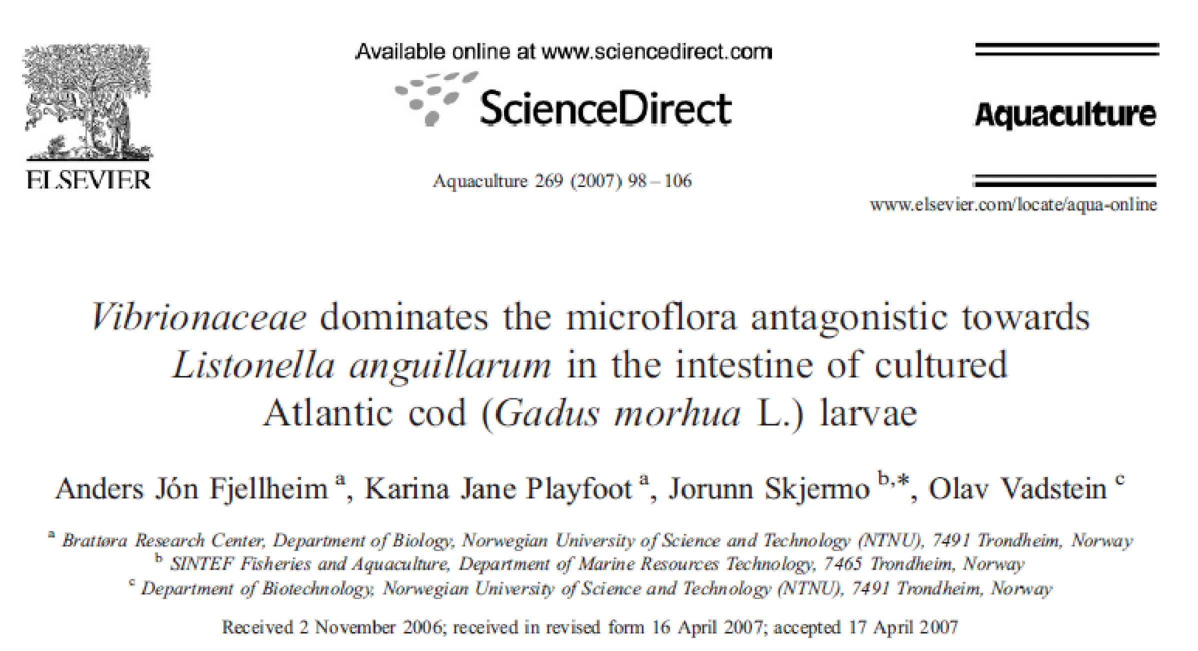
The PhylOgenetic Web Repeater (POWER) allows users performing phylogenetic analysis with molecular data by most programs of PHYLIP package repeatedly. POWER provide two pipelines to process the analysis. One of them includes multiple sequence alignment (MSA) at the begining of the pipeline whereas the other begin phylogenetic analysis with aligned sequence.

Please start your analysis by selecting the pipeline and the data type:

Pipeline	● MSA + Phylogenetic Analysis(Input the FASTA format) ○ Phylogenetic Analysis Only(Input the PHYLIP format)
Sequence Type	● DNA ○ Protein

http://power.nhri.org.tw/, in the Demo page

Conduct Distance Method in POWER



Available online at www.sciencedirect.com

ScienceDirect

Aquaculture

Aquaculture 269 (2007) 98 – 106

www.elsevier.com/locate/aqua-online

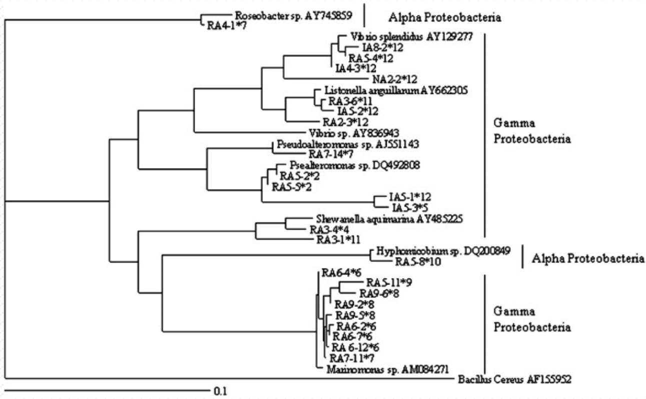
ELSEVIER

Vibrionaceae dominates the microflora antagonistic towards *Listonella anguillarum* in the intestine of cultured Atlantic cod (*Gadus morhua* L.) larvae

Anders Jón Fjellheim [a], Karina Jane Playfoot [a], Jorunn Skjermo [b,*], Olav Vadstein [c]

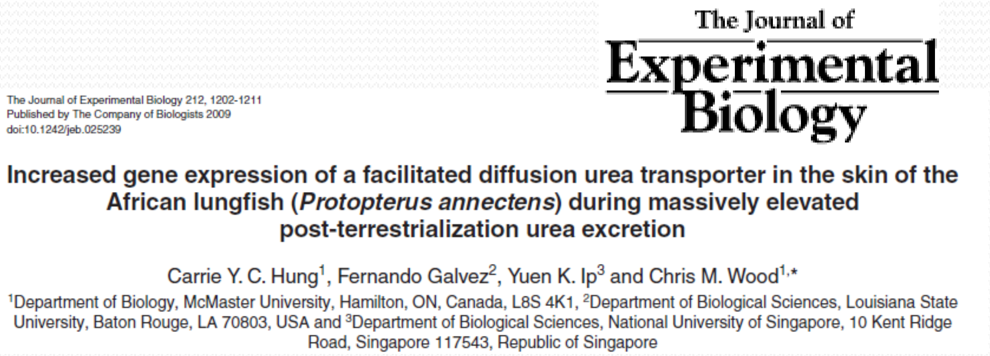
[a] *Brattøra Research Center, Department of Biology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*
[b] *SINTEF Fisheries and Aquaculture, Department of Marine Resources Technology, 7465 Trondheim, Norway*
[c] *Department of Biotechnology, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*

Received 2 November 2006; received in revised form 16 April 2007; accepted 17 April 2007



The DNA sequences were aligned to known sequences in the GenBank database using BLAST (Altschul et al., 1990). Phylogenetic relationships were inferred using the neighbour joining method (NJ), based on the Kimura two-parameter model (K2P), in **the PhylOgenetic WEb Repeater (POWER)** (Lin et al., 2005).

NJ with 1000 Replicates in POWER



The Journal of Experimental Biology 212, 1202-1211
Published by The Company of Biologists 2009
doi:10.1242/jeb.025239

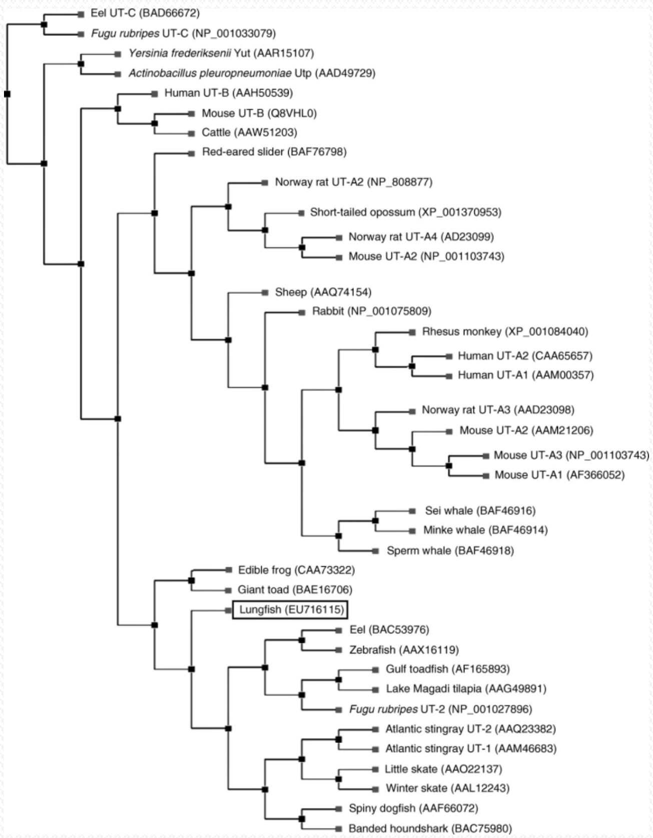
The Journal of Experimental Biology

Increased gene expression of a facilitated diffusion urea transporter in the skin of the African lungfish (*Protopterus annectens*) during massively elevated post-terrestrialization urea excretion

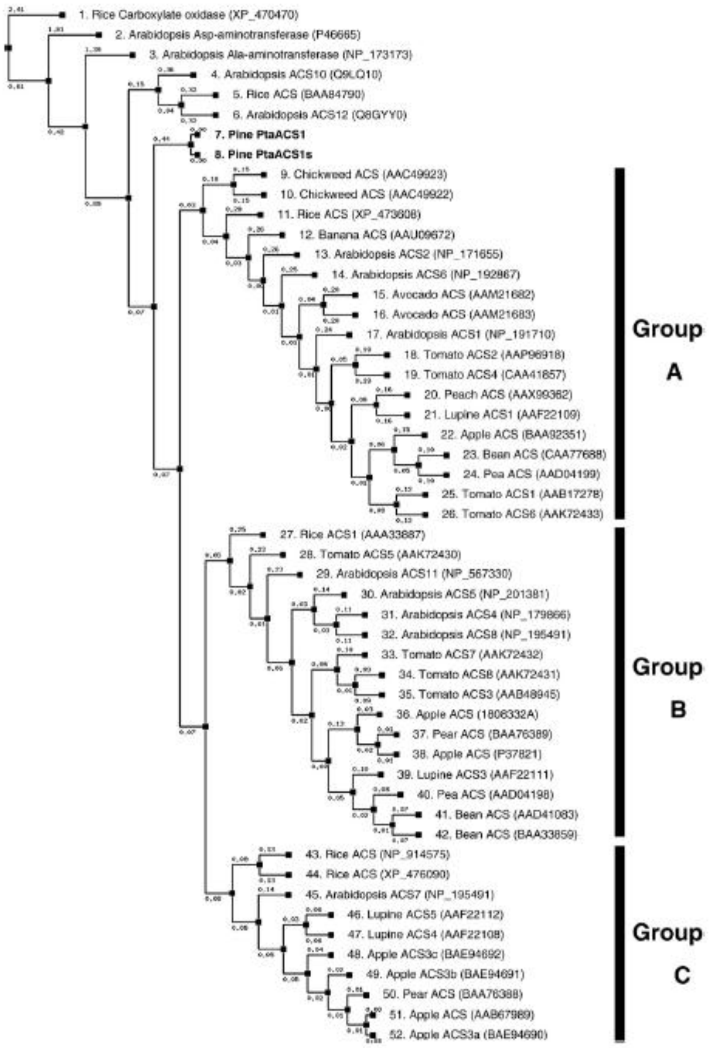
Carrie Y. C. Hung[1], Fernando Galvez[2], Yuen K. Ip[3] and Chris M. Wood[1,*]

[1]Department of Biology, McMaster University, Hamilton, ON, Canada, L8S 4K1, [2]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA and [3]Department of Biological Sciences, National University of Singapore, 10 Kent Ridge Road, Singapore 117543, Republic of Singapore

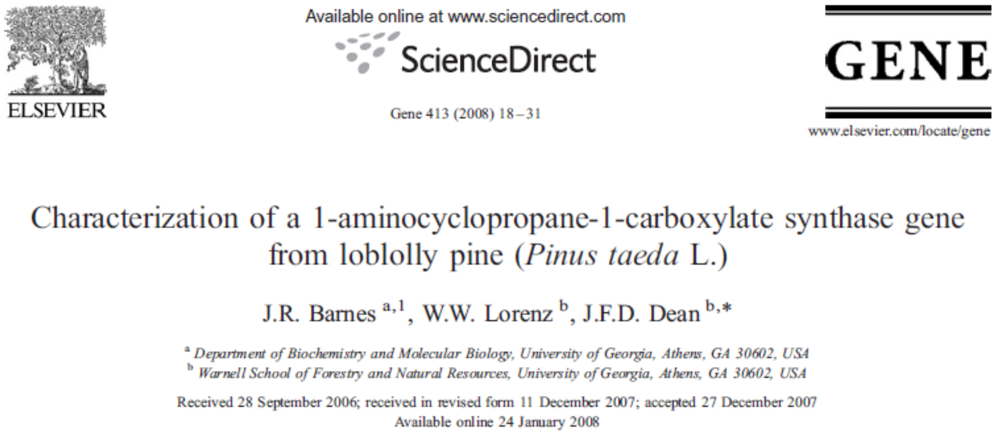
Phylogenetic tree of urea transporters across taxa. The protein sequences were aligned using ClustalW software, followed by neighbour-joining (NJ) matrix for tree reconstruction and evaluated by means of a bootstrap of 1000 replicates at http://power.nhri.org.tw


Eel UT-C (BAD66672)
Fugu rubripes UT-C (NP_001033079)
Yersinia frederiksenii Yut (AAR15107)
Actinobacillus pleuropneumoniae Utp (AAD49729)
Human UT-B (AAH50539)
Mouse UT-B (Q8VHL0)
Cattle (AAW51203)
Red-eared slider (BAF76798)
Norway rat UT-A2 (NP_808877)
Short-tailed opossum (XP_001370953)
Norway rat UT-A4 (AD23099)
Mouse UT-A2 (NP_001103743)
Sheep (AAQ74154)
Rabbit (NP_001075809)
Rhesus monkey (XP_001084040)
Human UT-A2 (CAA65657)
Human UT-A1 (AAM00357)
Norway rat UT-A3 (AAD23098)
Mouse UT-A2 (AAM21206)
Mouse UT-A3 (NP_001103743)
Mouse UT-A1 (AF366052)
Sei whale (BAF46916)
Minke whale (BAF46914)
Sperm whale (BAF46918)
Edible frog (CAA73322)
Giant toad (BAE16706)
Lungfish (EU716115)
Eel (BAC53976)
Zebrafish (AAX16119)
Gulf toadfish (AF165893)
Lake Magadi tilapia (AAG49891)
Fugu rubripes UT-2 (NP_001027896)
Atlantic stingray UT-2 (AAQ23382)
Atlantic stingray UT-1 (AAM46683)
Little skate (AAO22137)
Winter skate (AAL12243)
Spiny dogfish (AAF66072)
Banded houndshark (BAC75980)

Perform ML in POWER



1. Rice Carboxylate oxidase (XP_470470)
2. Arabidopsis Asp-aminotransferase (P46665)
3. Arabidopsis Ala-aminotransferase (NP_173173)
4. Arabidopsis ACS10 (Q9LQ10)
5. Rice ACS (BAA84790)
6. Arabidopsis ACS12 (Q8GYY0)
7. Pine PtaACS1
8. Pine PtaACS1s
9. Chickweed ACS (AAC49923)
10. Chickweed ACS (AAC49922)
11. Rice ACS (XP_473608)
12. Banana ACS (AAU09672)
13. Arabidopsis ACS2 (NP_171655)
14. Arabidopsis ACS6 (NP_192867)
15. Avocado ACS (AAM21682)
16. Avocado ACS (AAM21683)
17. Arabidopsis ACS1 (NP_191710)
18. Tomato ACS2 (AAP96918)
19. Tomato ACS4 (CAA41857)
20. Peach ACS (AAX99362)
21. Lupine ACS1 (AAF22109)
22. Apple ACS (BAA92351)
23. Bean ACS (CAA77688)
24. Pea ACS (AAD04199)
25. Tomato ACS1 (AAB17278)
26. Tomato ACS6 (AAK72433)
27. Rice ACS1 (AAA33887)
28. Tomato ACS5 (AAK72430)
29. Arabidopsis ACS11 (NP_567330)
30. Arabidopsis ACS5 (NP_201381)
31. Arabidopsis ACS4 (NP_179866)
32. Arabidopsis ACS8 (NP_195491)
33. Tomato ACS7 (AAK72432)
34. Tomato ACS8 (AAK72431)
35. Tomato ACS3 (AAB48945)
36. Apple ACS (1806332A)
37. Pear ACS (BAA76389)
38. Apple ACS (P37821)
39. Lupine ACS3 (AAF22111)
40. Pea ACS (AAD04198)
41. Bean ACS (AAD41083)
42. Bean ACS (BAA33859)
43. Rice ACS (NP_914575)
44. Rice ACS (XP_476090)
45. Arabidopsis ACS7 (NP_195491)
46. Lupine ACS5 (AAF22112)
47. Lupine ACS4 (AAF22108)
48. Apple ACS3c (BAE94692)
49. Apple ACS3b (BAE94691)
50. Pear ACS (BAA76388)
51. Apple ACS (AAB67989)
52. Apple ACS3a (BAE94690)
Group A
Group B
Group C



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Gene 413 (2008) 18–31

GENE

www.elsevier.com/locate/gene

Characterization of a 1-aminocyclopropane-1-carboxylate synthase gene from loblolly pine (*Pinus taeda* L.)

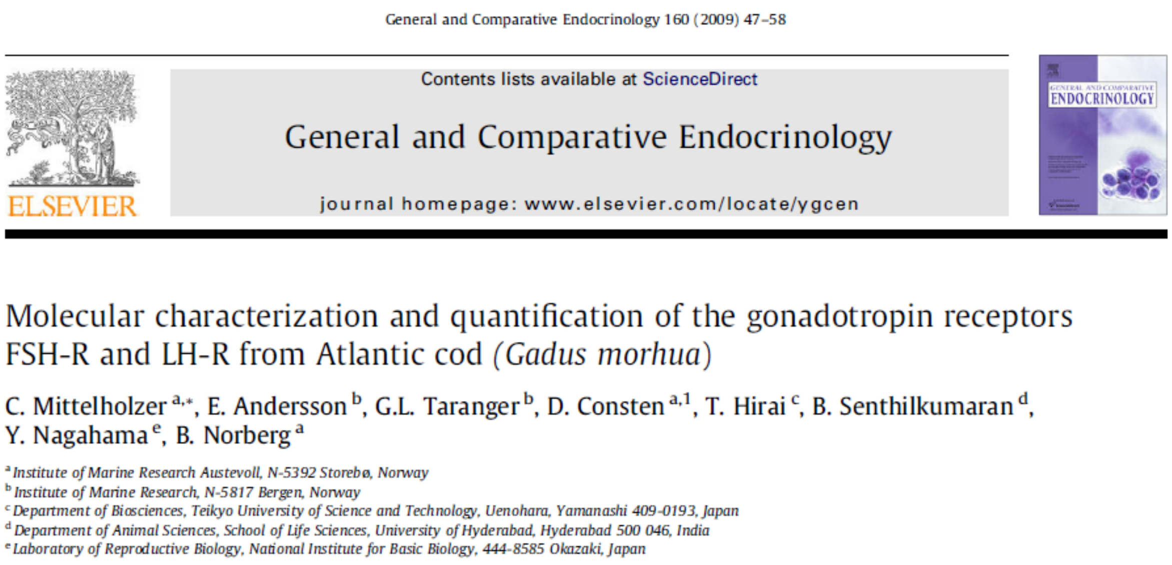
J.R. Barnes [a,1], W.W. Lorenz [b], J.F.D. Dean [b,*]

[a] *Department of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602, USA*
[b] *Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA*

Received 28 September 2006; received in revised form 11 December 2007; accepted 27 December 2007
Available online 24 January 2008

Phylogenetic tree depicting the relatedness of ACC synthase and aminotransferase protein sequences in GenBank.The phylogenetic tree was generated using the **POWER server (http://power.nhri.org.tw/)** with default parameters for the maximum likelihood method (ML) and molecular clock, but without bootstrapping or random input of sequences.

Execute MP in POWER



General and Comparative Endocrinology 160 (2009) 47–58

Contents lists available at ScienceDirect

General and Comparative Endocrinology

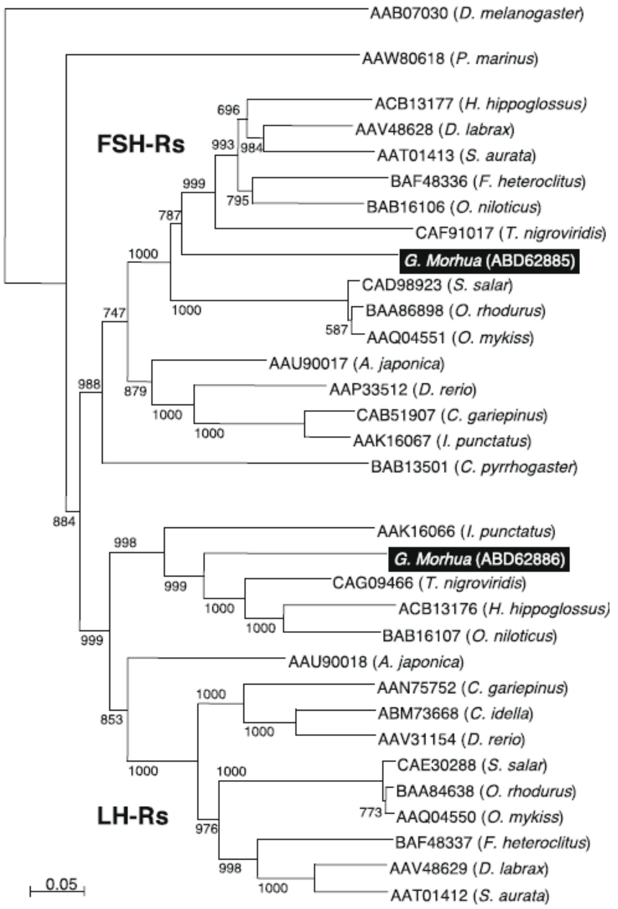
journal homepage: www.elsevier.com/locate/ygcen

Molecular characterization and quantification of the gonadotropin receptors FSH-R and LH-R from Atlantic cod (*Gadus morhua*)

C. Mittelholzer[a,*], E. Andersson[b], G.L. Taranger[b], D. Consten[a,1], T. Hirai[c], B. Senthilkumaran[d], Y. Nagahama[e], B. Norberg[a]

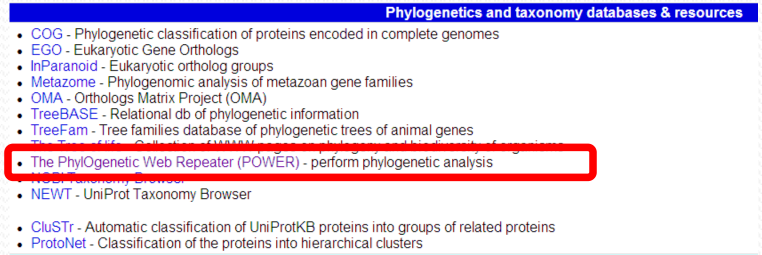
[a] Institute of Marine Research Austevoll, N-5392 Storebø, Norway
[b] Institute of Marine Research, N-5817 Bergen, Norway
[c] Department of Biosciences, Teikyo University of Science and Technology, Uenohara, Yamanashi 409-0193, Japan
[d] Department of Animal Sciences, School of Life Sciences, University of Hyderabad, Hyderabad 500 046, India
[e] Laboratory of Reproductive Biology, National Institute for Basic Biology, 444-8585 Okazaki, Japan

Phylogenetic comparison of fish full-length FSH-R and LH-R amino acid sequences analysed by **POWER** using **maximum parsimony (MP)** and default settings. A rooted consensus phylogenetic tree generated by means of the Neighbor-Joining algorithm, using the LGR sequence of the fruit fly (Drosophila melanogaster) and sea lamprey (Petromyzon marinus) as outgroups was drawn with njplot. Bootstrap values from 1000 replicates are indicated for each tree node.



POWER Listed in

- PHYLIP Programs maintained by Joe Felsenstein
  - Recent listings:
    - POWER server (26 August 2007) to align sequences and infer phylogenies, http://evolution.genetics.washington.edu/phylip/software.serv.html
- BioToolKit by CSHL press (BioSupplynet.com)
  - ALL CATEGORIES / GENOMICS RESOURCES / **EVOLUTIONARY AND COMPARATIVE BIOLOGY (80)**
- Bioinformatics Links Directory
  - DNA : Phylogeny Reconstruction
- ONLINE ANALYSIS TOOLS (http://molbiol-tools.ca/)
- ExPASy (Phylogenetics and taxonomy databases & resources)



Phylogenetics and taxonomy databases & resources

- COG - Phylogenetic classification of proteins encoded in complete genomes
- EGO - Eukaryotic Gene Orthologs
- InParanoid - Eukaryotic ortholog groups
- Metazome - Phylogenomic analysis of metazoan gene families
- OMA - Orthologs Matrix Project (OMA)
- TreeBASE - Relational db of phylogenetic information
- TreeFam - Tree families database of phylogenetic trees of animal genes
- The Tree of life - Collection of WWW pages on phylogeny and biodiversity of organisms
- The PhylOgenetic Web Repeater (POWER) - perform phylogenetic analysis
- NCBI Taxonomy Browser
- NEWT - UniProt Taxonomy Browser

- CluSTr - Automatic classification of UniProtKB proteins into groups of related proteins
- ProtoNet - Classification of the proteins into hierarchical clusters

Distance Method, MP and ML

- Which method should we choose?
- The main disadvantage of distance-matrix methods is their inability to efficiently use information about local high-variation regions that appear across multiple subtrees.
- ML is broadly similar to the maximum-parsimony (MP) method, but maximum likelihood allows additional statistical flexibility by permitting varying rates of evolution across both lineages and sites.
- ML, a better choice?

Maximum Likelihood

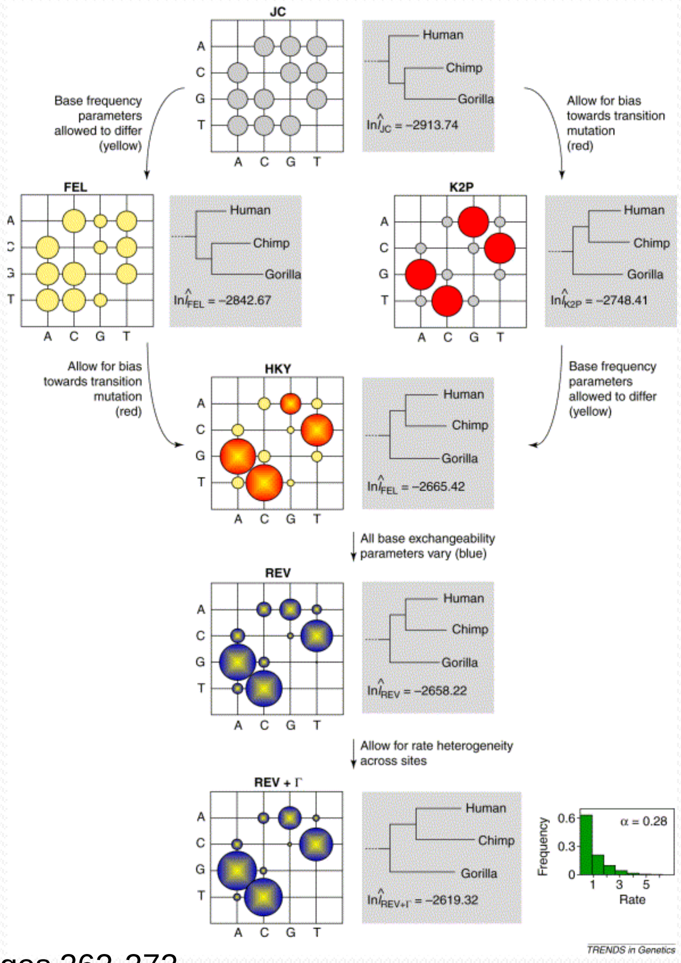
- Conditional probability of the data (Aligned sequences) given a hypothesis (a model of substitution with a set of parameter θ, and the tree τ, including topology and branch lengths)

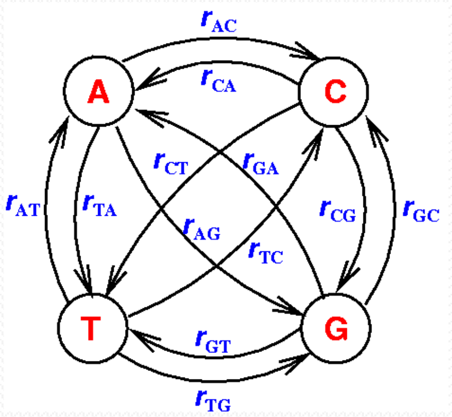
$$L(\tau, \theta) = \mathrm{Prob}(\mathrm{Data} \mid \tau, \theta)$$

Or

Prob(Aligned Sequences| tree, model of evolution)

Relationships among some standard models of nucleotide evolution

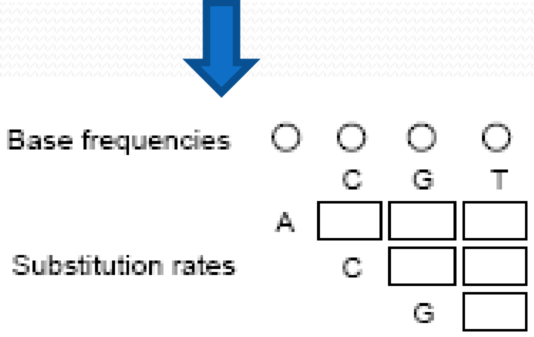

JC
A
C
G
T
A C G T
Human
Chimp
Gorilla
ln$\hat{l}_{JC}$ = −2913.74
Base frequency parameters allowed to differ (yellow)
Allow for bias towards transition mutation (red)
FEL
ln$\hat{l}_{FEL}$ = −2842.67
K2P
ln$\hat{l}_{K2P}$ = −2748.41
Allow for bias towards transition mutation (red)
Base frequency parameters allowed to differ (yellow)
HKY
ln$\hat{l}_{FEL}$ = −2665.42
All base exchangeability parameters vary (blue)
REV
ln$\hat{l}_{REV}$ = −2658.22
Allow for rate heterogeneity across sites
REV + Γ
ln$\hat{l}_{REV+Γ}$ = −2619.32
Frequency
0.6
0.3
0
α = 0.28
1 3 5
Rate
TRENDS in Genetics



Trends in Genetics
Volume 17, Issue 5, 1 May 2001, Pages 262-272

Illustration of DNA Substitution Model

$$Q = \begin{pmatrix} -(x_1 + x_2 + x_3) & x_1 & x_2 & x_3 \\ \frac{\pi_1 x_1}{\pi_2} & -(\frac{\pi_1 x_1}{\pi_2} + x_4 + x_5) & x_4 & x_5 \\ \frac{\pi_1 x_2}{\pi_3} & \frac{\pi_2 x_4}{\pi_3} & -(\frac{\pi_1 x_2}{\pi_3} + \frac{\pi_2 x_4}{\pi_3} + x_6) & x_6 \\ \frac{\pi_1 x_3}{\pi_4} & \frac{\pi_2 x_5}{\pi_4} & \frac{\pi_3 x_6}{\pi_4} & -(\frac{\pi_1 x_3}{\pi_4} + \frac{\pi_2 x_5}{\pi_4} + \frac{\pi_3 x_6}{\pi_4}) \end{pmatrix}$$



Base frequencies ○ ○ ○ ○

Substitution rates

	C	G	T
A	□	□	□
C		□	□
G			□

GTR (for four characters, as is often the case in phylogenetics) requires 6 substitution rate parameters (x1~x6), as well as 4 equilibrium base frequency parameters.

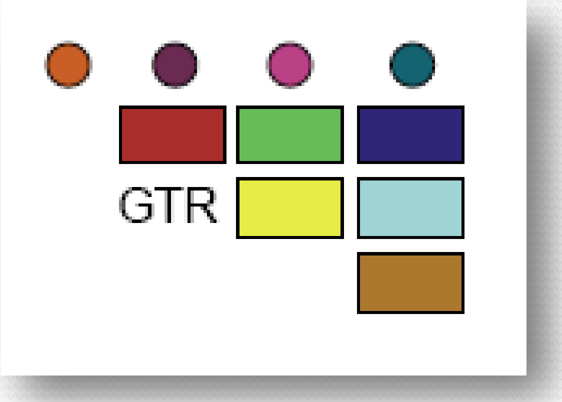
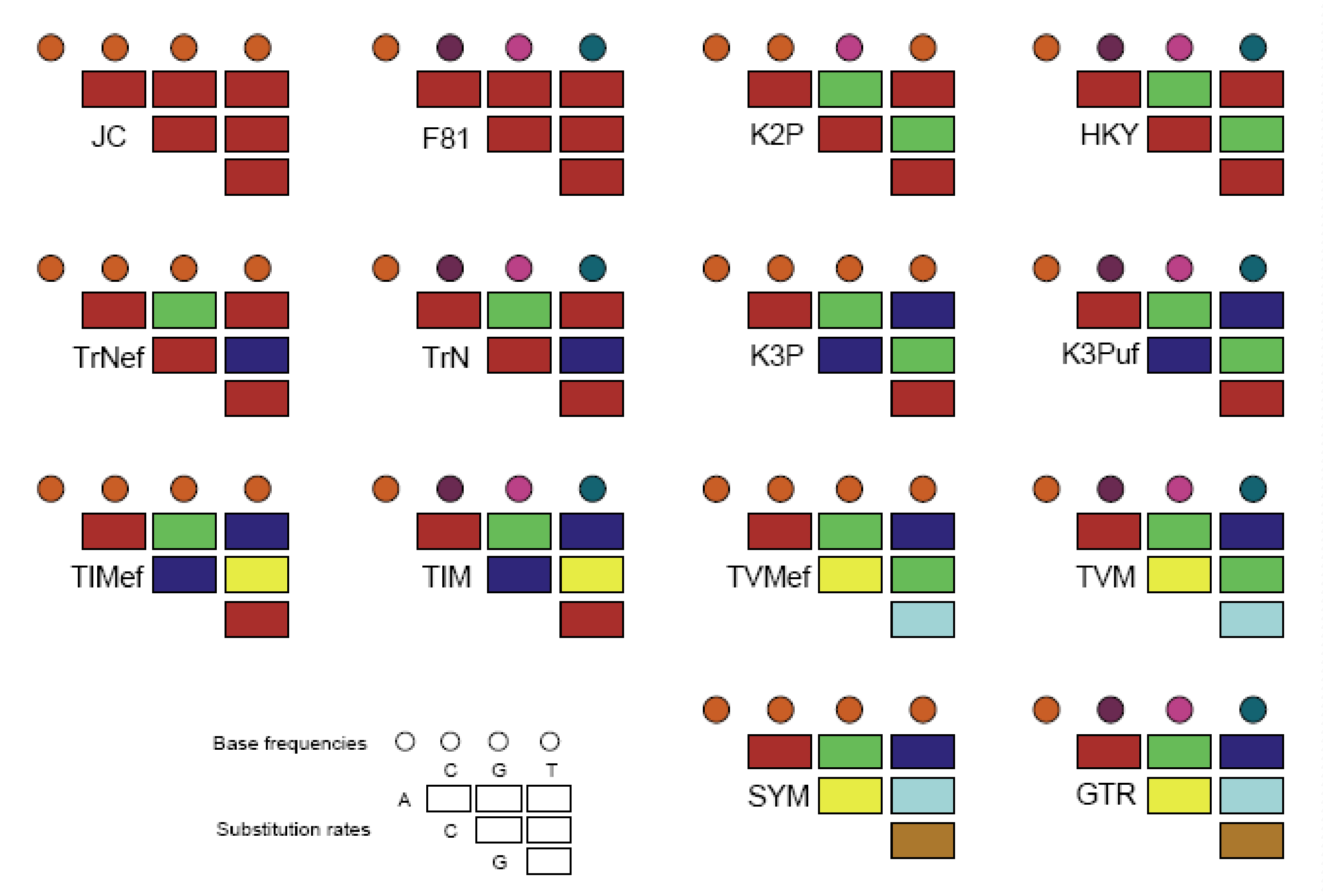
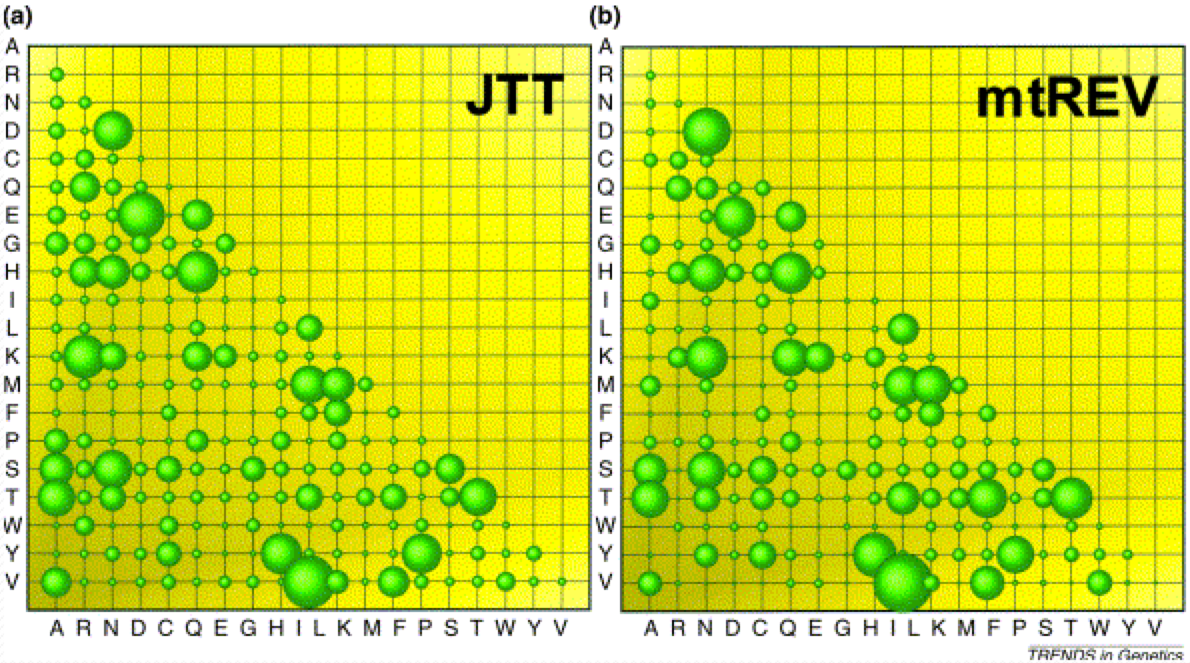


Illustration of Models for DNA



Models of Amino Acid Replacement



Trends in Genetics
Volume 17, Issue 5, 1 May 2001, Pages 262-272

Phylogenetic Reconstruction by Automatic Likelihood Model Selector (PALM) :

A Framework for Phylogenetic Analysis with the Best Substitution Model

陳淑華
sophia@iis.sinica.edu.tw

PALM
Phylogenetic reconstruction by Automatic Likelihood Model selector
Home
Demo
Help
Contact

PLoS ONE, 2009

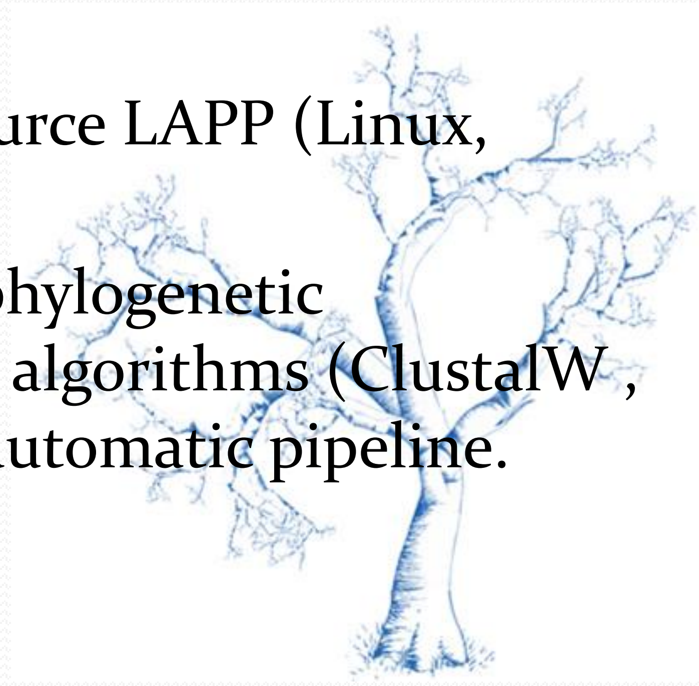
PLoS one
accelerating the publication of peer-reviewed science

Background for PALM

- Likelihood methods in phylogenetics relaxes the parameters for varying rates of evolution across both lineages and sites, which is robust in dealing with various extend of input sequence similarity.
- Model fitting has been suggested for many years, but many researchers select models arbitrarily. They often feel confusing either in making choice among models, or in dealing with the conflict on the results concluded by different models.
- The computing of likelihood method is intensive. Thus the ML-based model selecting procedure is hard to implement.
- Here we present the way to identify the best-fit model based on liklihood measurement. Consequently, model fitting is possible to be a routine practice integrated in a phylogenetic analysis.

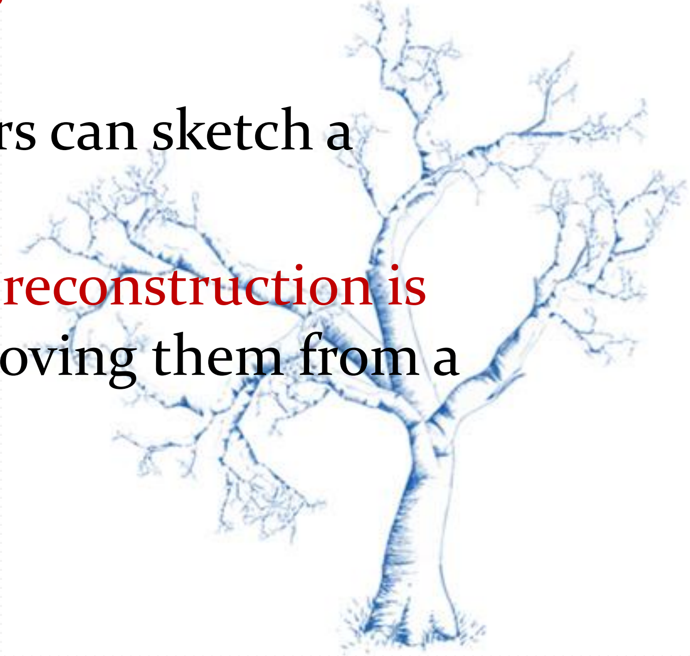
Motivation I

- Provide a seamless way to conduct the complicated phylogenetic analysis for biologists and biomedical researchers.
- An integrated and user-friendly framework for conducting molecular phylogenetic analysis
- PALM is constructed on an open-source LAPP (Linux, Apache, PostgreSql, PHP) structure
- PALM infers genetic distances and phylogenetic relationships using well-established algorithms (ClustalW , PhyML, ProtTest, Modeltest) in an automatic pipeline.



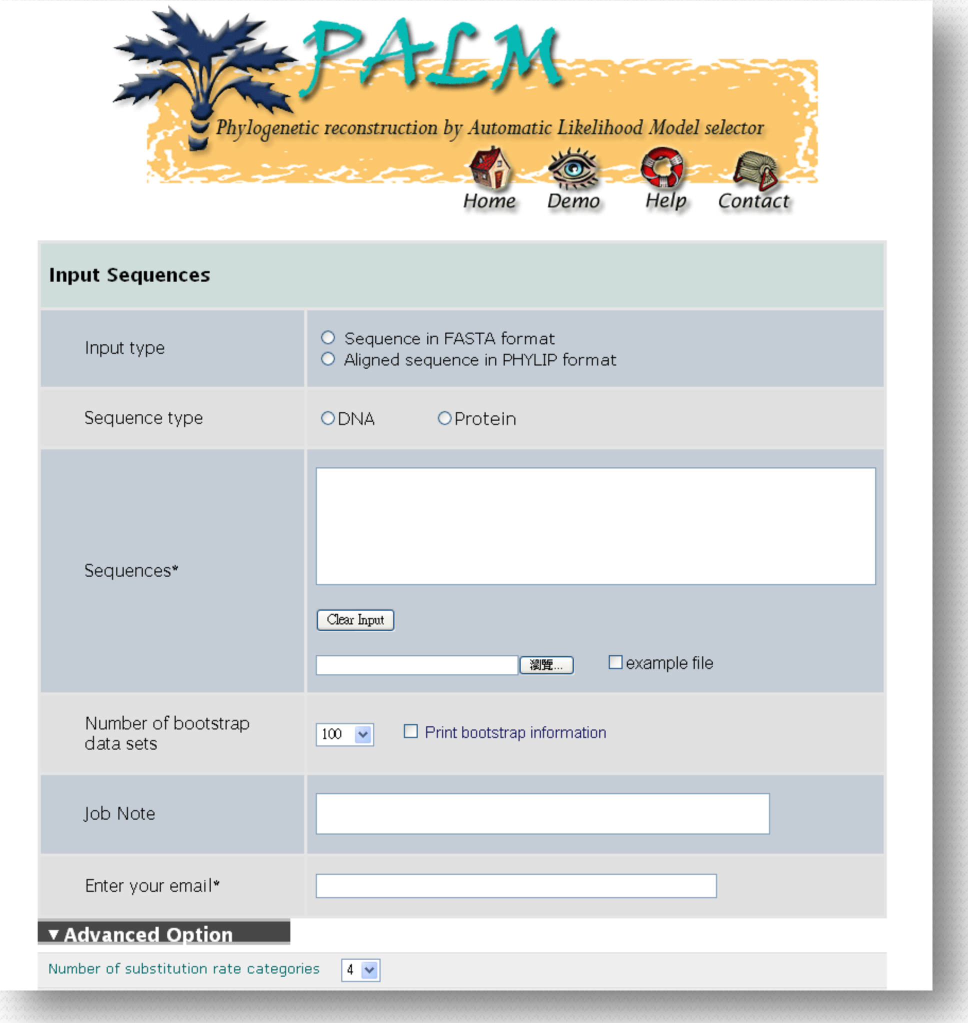
Motivation II

- Fitness of model can be measured and selected by following criteria: likelihood ratio tests (hLRTs), Akaike information criterion (AIC), and Bayesian information criterion (BIC)
- PALM helps user to construct the phylogenetic relationship by ML-based method with bootstrap using the best-fit substitution model.
- Through the friendly web interface, users can sketch a phylogenetic tree effortlessly
- Furthermore, iteration on phylogenetic reconstruction is possible by adding sequences to, or removing them from a previously result.



Component Programs of PALM

- PhyML 3.0
- ModelTest 3.7
- ProtTest 2.0
- ClustalW 2.0.3
- ReadSeq


PALM
Phylogenetic reconstruction by Automatic Likelihood Model selector
Home
Demo
Help
Contact

Input Sequences

Input type	○ Sequence in FASTA format ○ Aligned sequence in PHYLIP format
Sequence type	○DNA ○Protein
Sequences*	Clear Input 瀏覽... □example file
Number of bootstrap data sets	100 □ Print bootstrap information
Job Note	
Enter your email*	

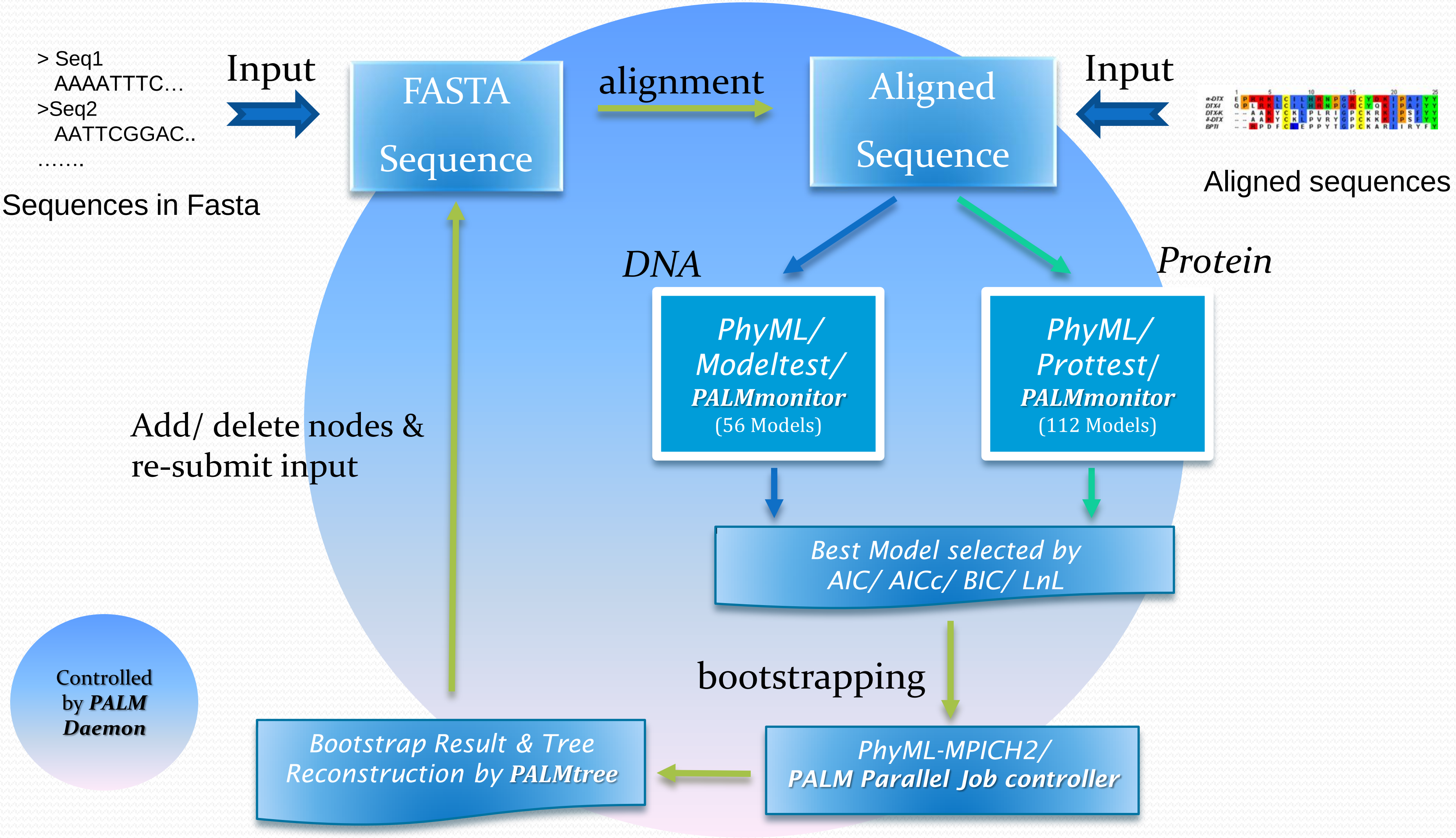
▾ Advanced Option

Number of substitution rate categories 4

Models Used in PALM

- For DNA (56 models)
  - JC69, K80, F81, HKY, TrN, TrNef, K3P, K3Puf, TIM, TIMef, TVM, TVMef, SYM, GTR
  - Options of +I, +G
- For Protein (112 models), **Time consuming**
  - LG, DCMut, JTT, MtREV, MtMam, MtArt, Dayhoff, WAG, RtREV, CpREV, Blosum62, VT, HIVb, HIVw
  - Options of +I, +G, +F

Flowchart of PALM



> Seq1
AAAATTTC…
>Seq2
AATTCGGAC..
…….

Sequences in Fasta

Input

FASTA Sequence

alignment

Aligned Sequence

Input

Aligned sequences

DNA

Protein

PhyML/ Modeltest/ ***PALMmonitor*** (56 Models)

PhyML/ Prottest/ ***PALMmonitor*** (112 Models)

Best Model selected by AIC/ AICc/ BIC/ LnL

bootstrapping

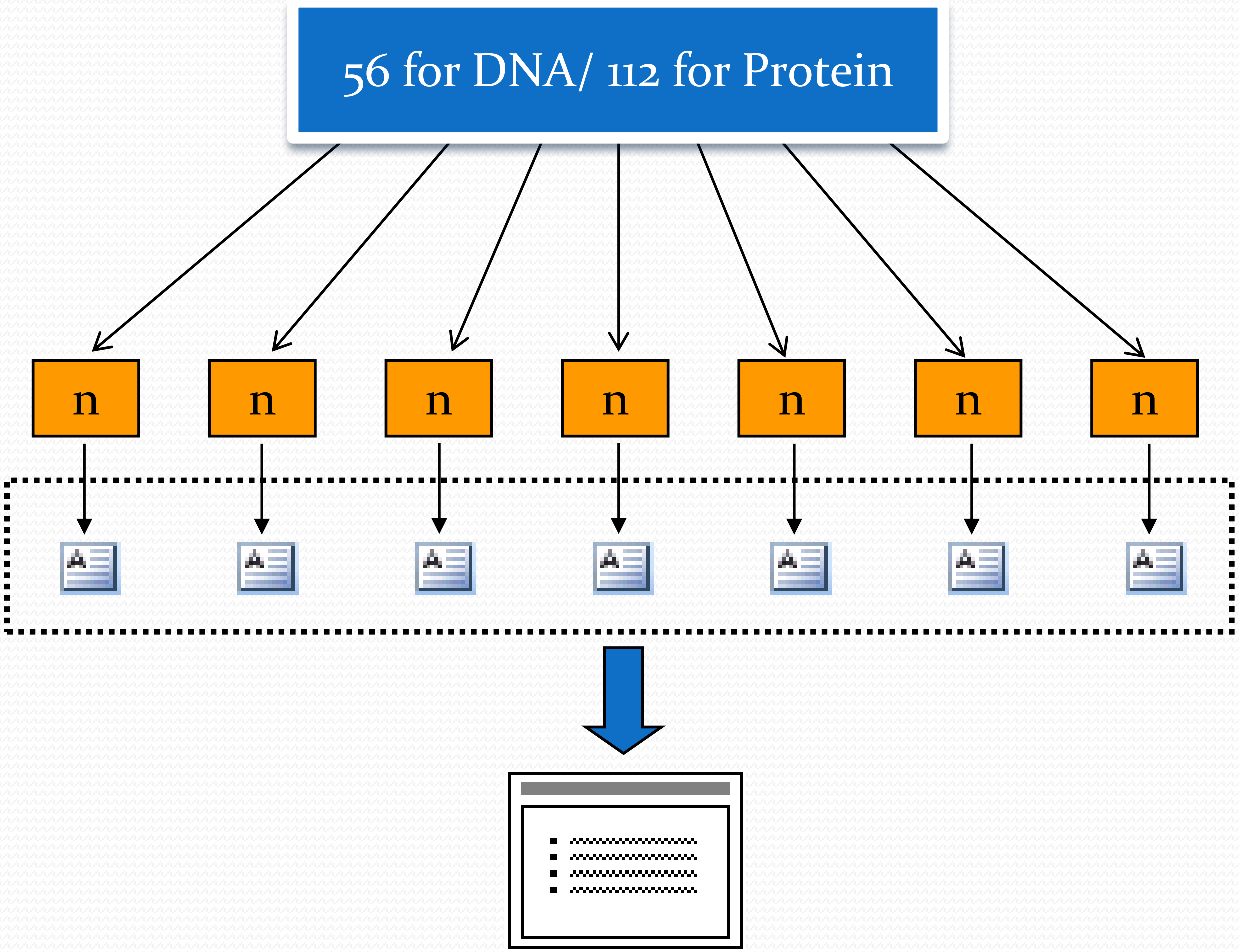
PhyML-MPICH2/ ***PALM Parallel Job controller***

Bootstrap Result & Tree Reconstruction by ***PALMtree***

Add/ delete nodes & re-submit input

Controlled by ***PALM Daemon***

Distribution Computing by PalmMonitor for the Likelihood Estimation of Models



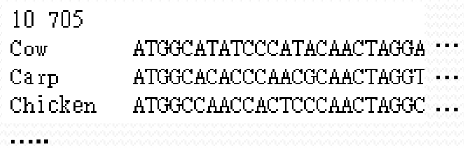
56 for DNA/ 112 for Protein

7 parallel processes

Input and Output of PALM

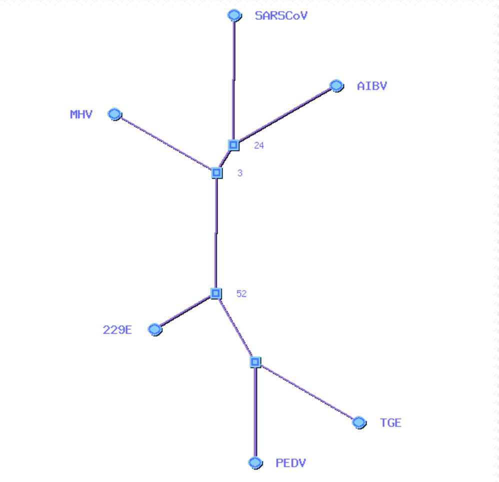
- Input format (Protein and DNA)
  - FASTA format
  - Phylip format: Aligned Sequences
  - User tree (if a valid tree is submitted)

```
> Seq1
  AAAATTTC…
>Seq2
  AATTCGGAC...
```



```
10 705
Cow        ATGGCATATCCCATACAACTAGGA ···
Carp       ATGGCACACCCAACGCAACTAGGT ···
Chicken    ATGGCCAACCACTCCCAACTAGGC ...
.....
```

- Output
  - Tree topology
  - Tree file in Newick format
  - Aligned sequence in phylip format
  - The best model selected by PALM
  - Likelihoods of all available models

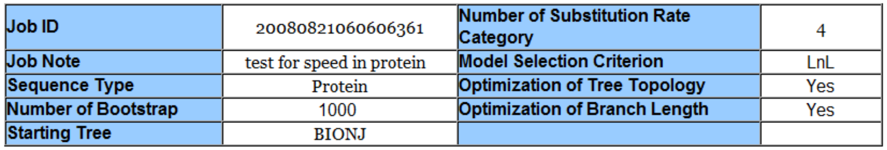


Result of PALM

A

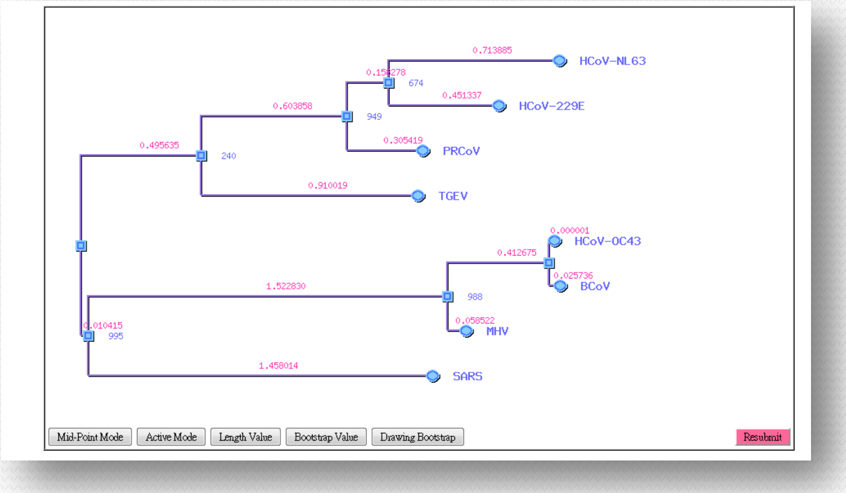
PALM
Phylogenetic reconstruction by Automatic Likelihood Model selector
Home
Demo
Help
Contact

PALM Result

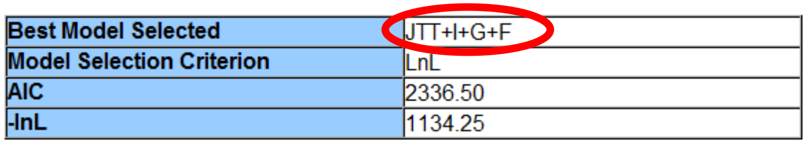


Job ID	20080821060606361	**Number of Substitution Rate Category**	4
Job Note	test for speed in protein	**Model Selection Criterion**	LnL
Sequence Type	Protein	**Optimization of Tree Topology**	Yes
Number of Bootstrap	1000	**Optimization of Branch Length**	Yes
Starting Tree	BIONJ		

B

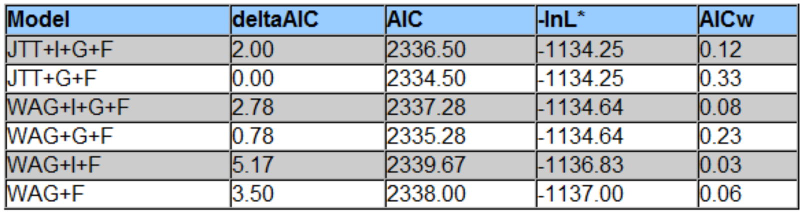


C



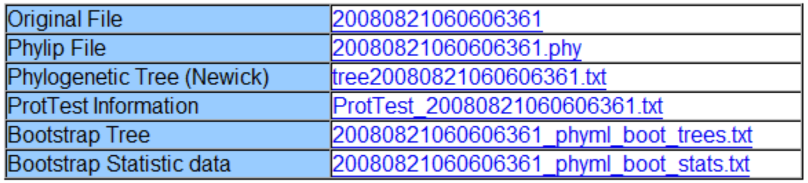
Best Model Selected	JTT+I+G+F
Model Selection Criterion	LnL
AIC	2336.50
-lnL	1134.25

D



Model	deltaAIC	AIC	-lnL*	AICw
JTT+I+G+F	2.00	2336.50	-1134.25	0.12
JTT+G+F	0.00	2334.50	-1134.25	0.33
WAG+I+G+F	2.78	2337.28	-1134.64	0.08
WAG+G+F	0.78	2335.28	-1134.64	0.23
WAG+I+F	5.17	2339.67	-1136.83	0.03
WAG+F	3.50	2338.00	-1137.00	0.06

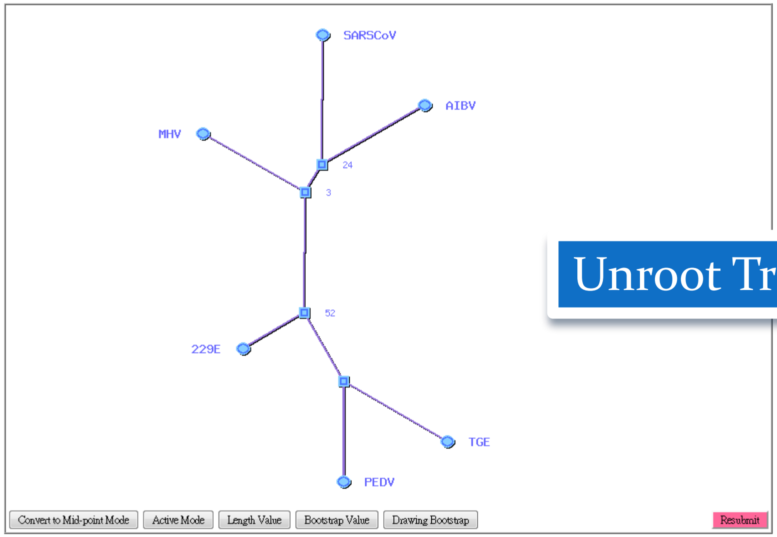
E



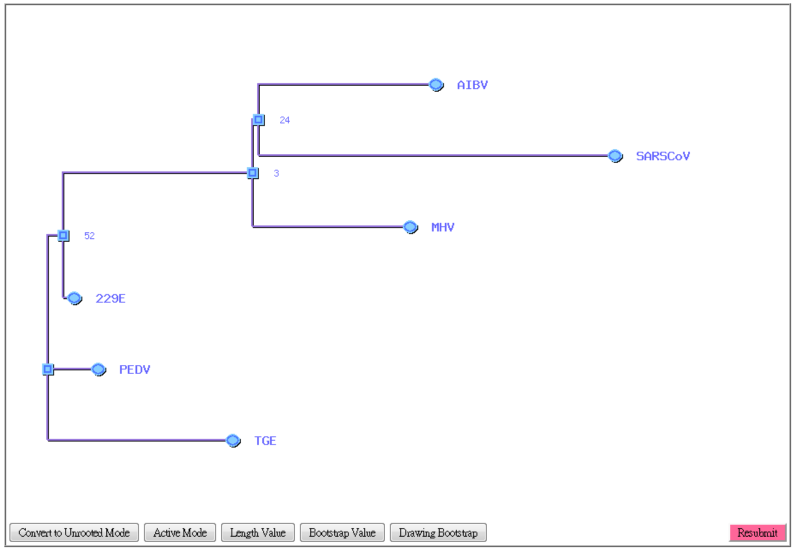
Original File	20080821060606361
Phylip File	20080821060606361.phy
Phylogenetic Tree (Newick)	tree20080821060606361.txt
ProtTest Information	ProtTest_20080821060606361.txt
Bootstrap Tree	20080821060606361_phyml_boot_trees.txt
Bootstrap Statistic data	20080821060606361_phyml_boot_stats.txt

The job is computed approximately in 47 minute(s).

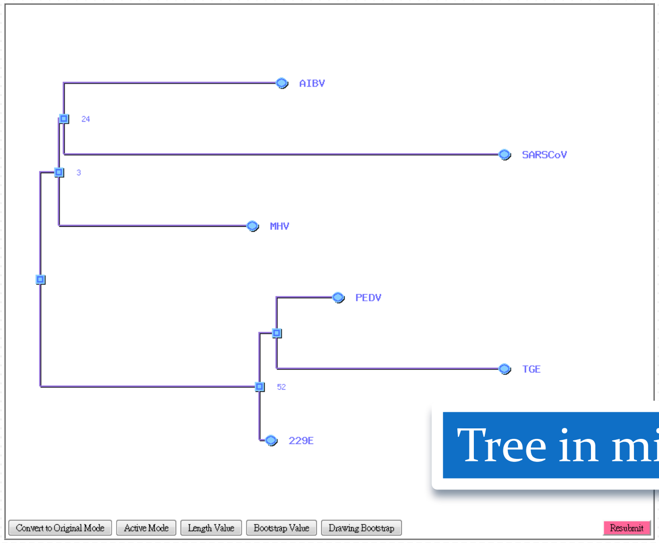
PALMtree



Unroot Tree

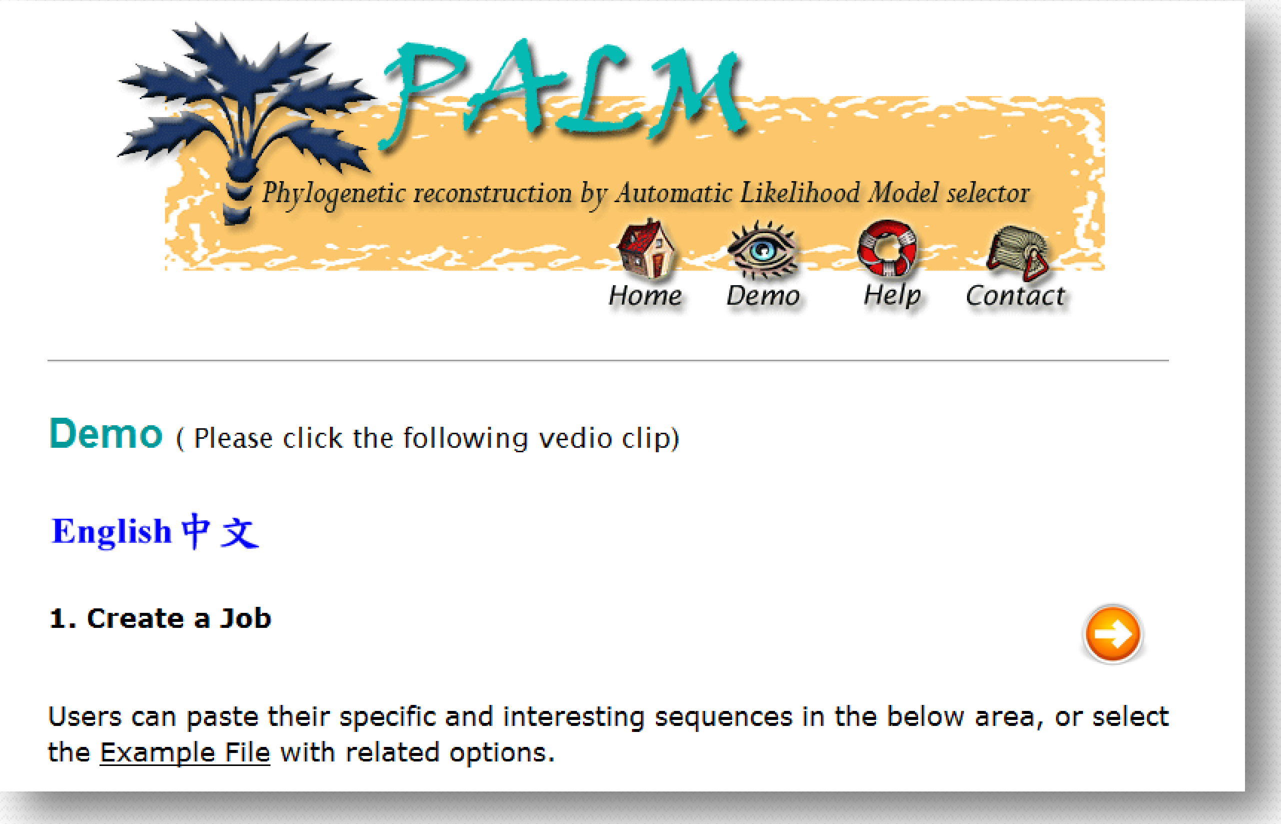


Original Tree



Tree in midpoint

Demo Flash of PALM



PALM
Phylogenetic reconstruction by Automatic Likelihood Model selector
Home
Demo
Help
Contact

Demo (Please click the following vedio clip)

English 中文

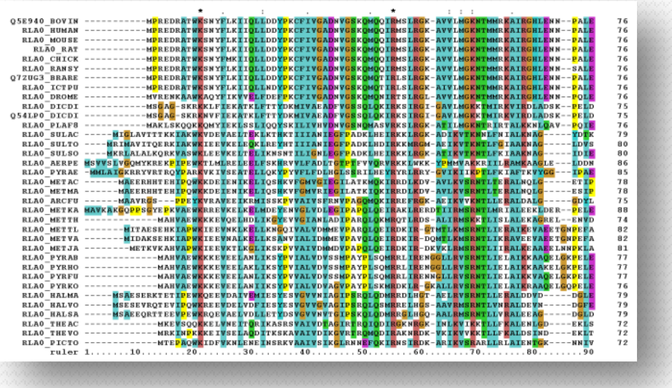
1. Create a Job

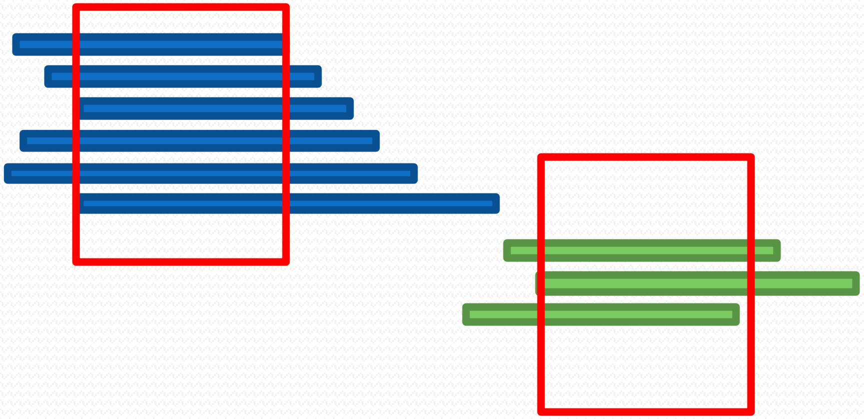
Users can paste their specific and interesting sequences in the below area, or select the Example File with related options.

http://palm.iis.sinica.edu.tw/demo.html

Some Suggestions

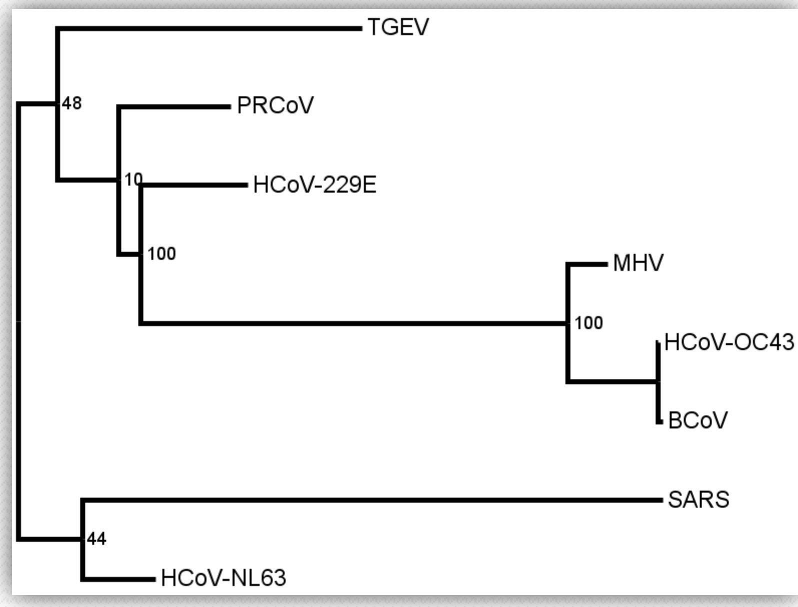
- Please be patient and make a reasonable choice of the input sequence set
- Only **well aligned** sequences lead to meaningful phylogenetic result.
- RNA editing may introduce bias during analysis. Avoid those regions that may have such conditions.





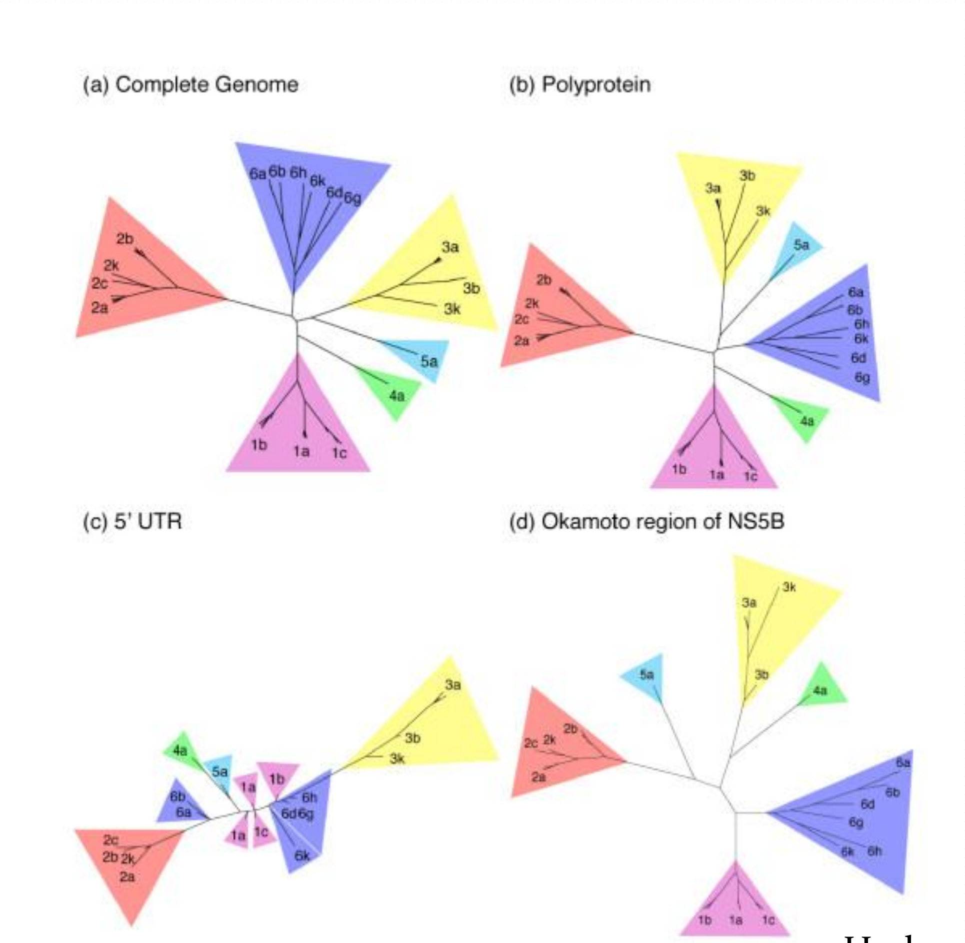
Bootstrap (BS) Analysis

- Bootstrap analysis is the most popular method for statistical evaluation of phylogenies.
- In general:
  - **BS >95%: Often close to 100% confidence in that branch**
  - **BS>75%: Often close to 95% confidence in that branch**
  - BS<75% : Maybe a correct clade, while the original bias cannot be corrected by the re-sampling process.



Input Sequences Make the Tree Different

HIV



Hraber *et al. Virology Journal* 2006 3:103 doi:10.1186/1743-422X-3-103

Future Plans for PALM

- Gateway to integrate users-defined substitution models
- Stand-a-lone version of *PALM*
- Improve and optimize the performance of whole pipeline by applying parallel computing/cloud computing
- Implement of advanced, sophisticated phylogenetic inference methods such as MrBayes.

Acknowledgement

國家衛生研究院
National Health Research Institutes

Chia-Ling Chen
Chieh-Hwa Lin
Li-Wei Lai
Shu-Juan Hsu
Ming-Hsin Tasi
Chao A. Hsiung

中央研究院
資訊科學研究所
Institute of Information Science
Academia Sinica

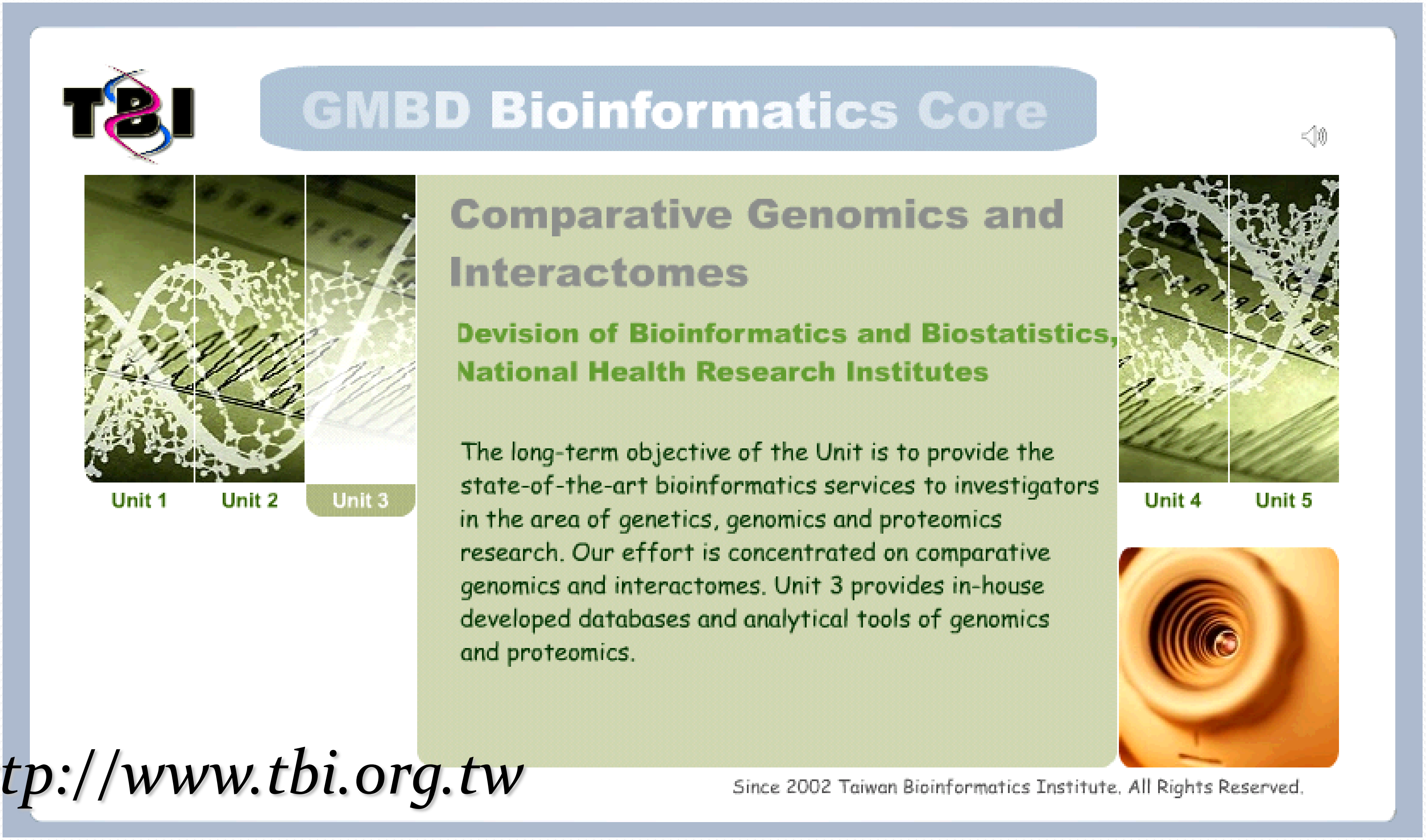
Daniel, Sheng-Yao, Su
Pan-Han Kuo
Tengi Huang
Chen-Zen Lo
Linda, Yi-Shuian Lu

中央研究院 資訊科技創新研究中心
Research Center for Information Technology Innovation

行政院國家科學委員會
National Science Council

基因體醫學國家型科技計畫
National Research Program
for Genomic Medicine

Bioinformatics Core for Genomic Medicine and Biotechnology Development



Comparative Genomics and Interactomes

Devision of Bioinformatics and Biostatistics, National Health Research Institutes

The long-term objective of the Unit is to provide the state-of-the-art bioinformatics services to investigators in the area of genetics, genomics and proteomics research. Our effort is concentrated on comparative genomics and interactomes. Unit 3 provides in-house developed databases and analytical tools of genomics and proteomics.

Since 2002 Taiwan Bioinformatics Institute. All Rights Reserved.

http://www.tbi.org.tw

Thanks for Your Attention!

And Time for Practice