Deciphering the Biological Problems in the Approach of Systems Biology

林 仲 彥, Chung-Yen Lin Ph.D.

Assistant Research Fellow

cylin@iis.sinica.edu.tw

Laboratory of Systems Biology and Network Biology

Institute of Information Science, Academia Sinica

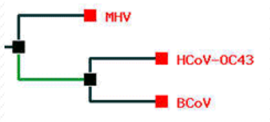
Aug 12, 2008

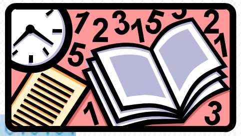
中央研究院
資訊科學研究所
Institute of Information Science
Academia Sinica

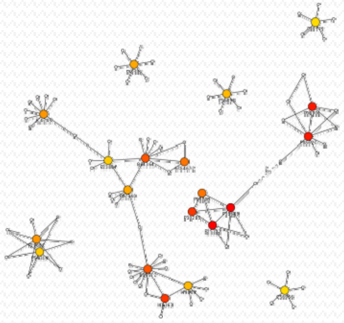
國家衛生研究院
National Health Research Institutes

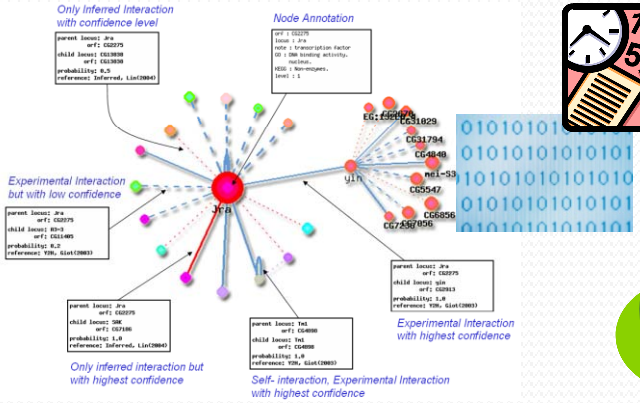
Outline

- **Solving the Biological problems in Computational methods and statistics**
  - Genomics studies for high throughput research
  - Phylogenetics analysis
  - Protein interactome
  - Network comparison and topological analysis
  - Ongoing projects









PDA
Primer Design Assistant

Platform Based on LAMP/ LAPP

Linux

Operation System

Apache (with OpenSSL)

Webserver

MySQL/**P**ostgrSQL

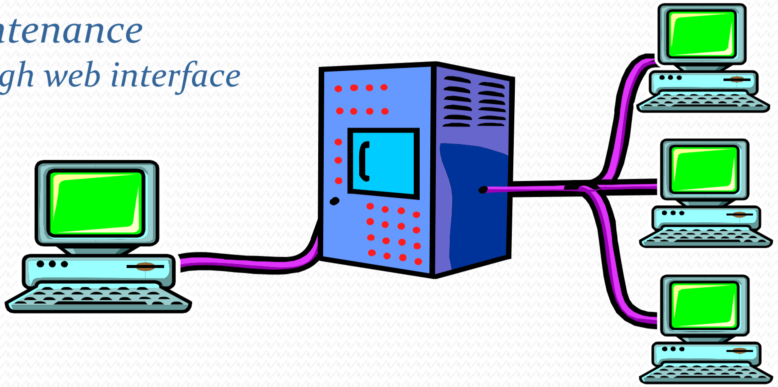
Relational Database

PHP

Server-side HTML embedded scripting language with GD library

Query by Web interface (Win9x/Me/2000/Mac/Unix /Linux/Solaris)

Maintenance through web interface



Web database
Linux
Apache webServer,
MySQL
PHP control language

Design of Primers and Probes
for High and low Throughput Research

- ✓Primer Design Assistant (PDA)
- ✓Unique Probe Selector (UPS)

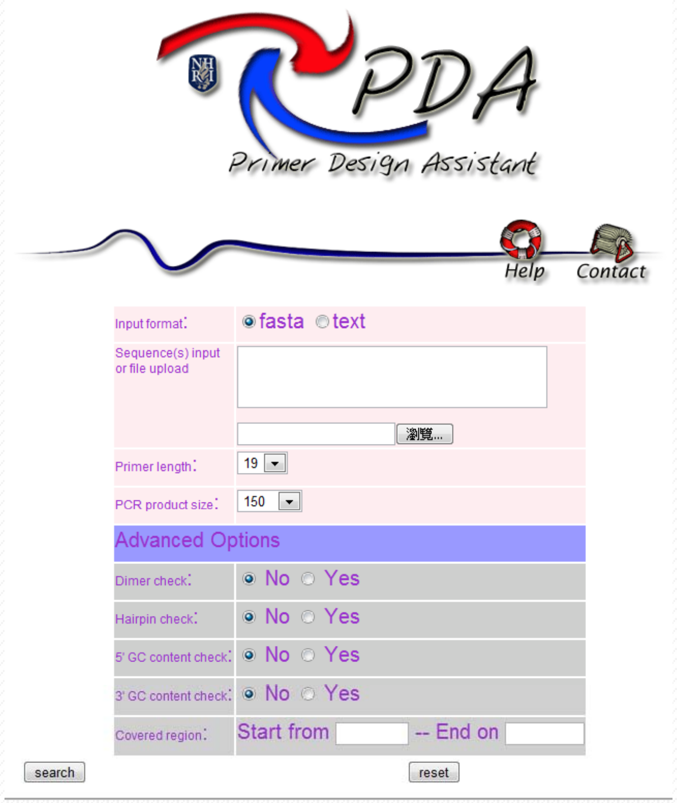
NHRI PDA
Primer Design Assistant

UPS
Unique Probe Selector

Motivation for PDA

- **Integration of experiences from wet-lab and computational technology** to perform primer design in large scale PCR under similar Tm value.
- Primer Design Assistant (PDA) is a **web interface primer design service combined with thermodynamic theory** to evaluate the fitness of primers.
- It runs in a Linux–Apache–MySQL–PHP structure on a PC equipped with dual CPU (Intel Pentium III 1.4 GHz) and 512 Mb of RAM.
- **A succinct user interface** of PDA is accomplished by built-in parameters setting. Advanced options on 5' GC content, 3' GC content, dimer check and hairpin check are available.
- **PDA accepts single sequence query or multiple ones in FASTA format**. It produces optimal and homogenous primer pairs that meet the need in experimental design with **large-scaled PCR** amplifications.

Genomics Studies For High Throughput Research : PDA



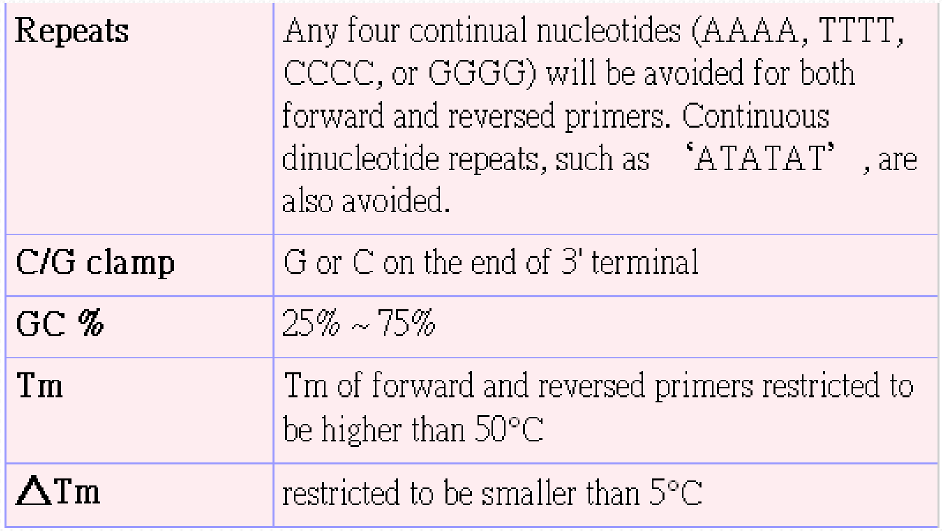
Input format:	● fasta ○ text
Sequence(s) input or file upload	瀏覽...
Primer length:	19
PCR product size:	150
Advanced Options	
Dimer check:	● No ○ Yes
Hairpin check:	● No ○ Yes
5' GC content check:	● No ○ Yes
3' GC content check:	● No ○ Yes
Covered region:	Start from -- End on

search reset

- ***Primers designed through PDA has been experimentally proved to reach 97% successful rate***
- ***PDA can be used to design the primers set for high through put experiments. For example , for 96 /384 format PCR Rx.***
- **http://dbb.nhri.org.tw/primer/**
- **Published on NAR 2003**

Criterion for PDA Setting

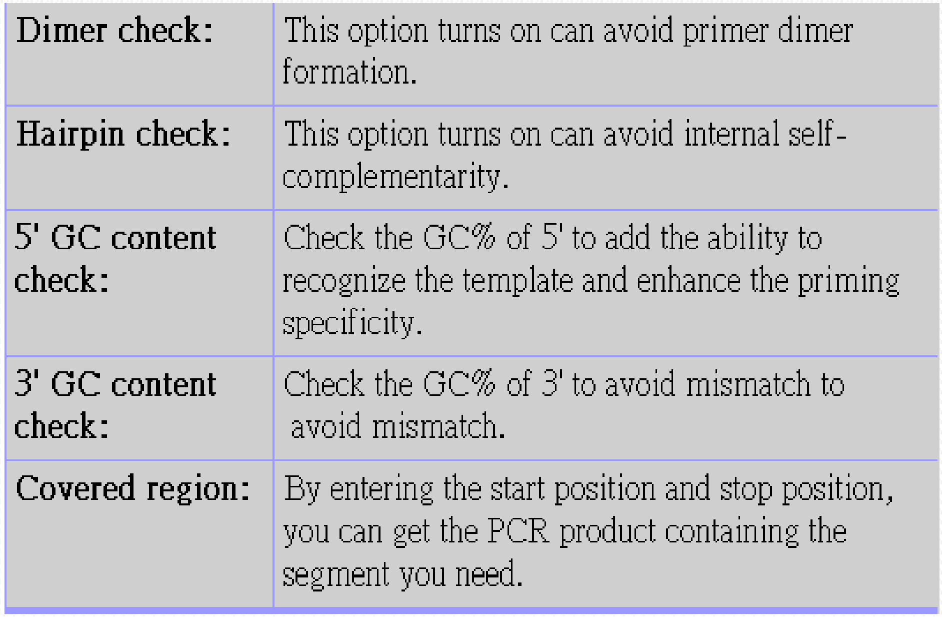
- Default Settings



Repeats	Any four continual nucleotides (AAAA, TTTT, CCCC, or GGGG) will be avoided for both forward and reversed primers. Continuous dinucleotide repeats, such as ‘ATATAT’ , are also avoided.
C/G clamp	G or C on the end of 3' terminal
GC %	25% ~ 75%
Tm	Tm of forward and reversed primers restricted to be higher than 50°C
ΔTm	restricted to be smaller than 5°C

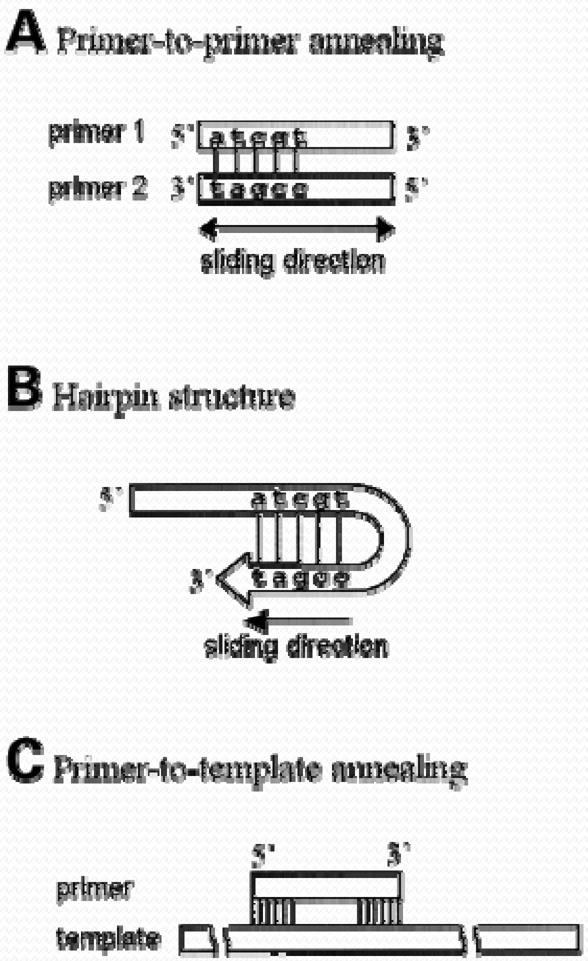
$$T_{\mathrm{m}}(^{\circ}\mathrm{C}) = 59.9 + (0.41 \times \mathrm{GC\ content}) - \left(\frac{675}{\mathrm{primer\ length}}\right)$$

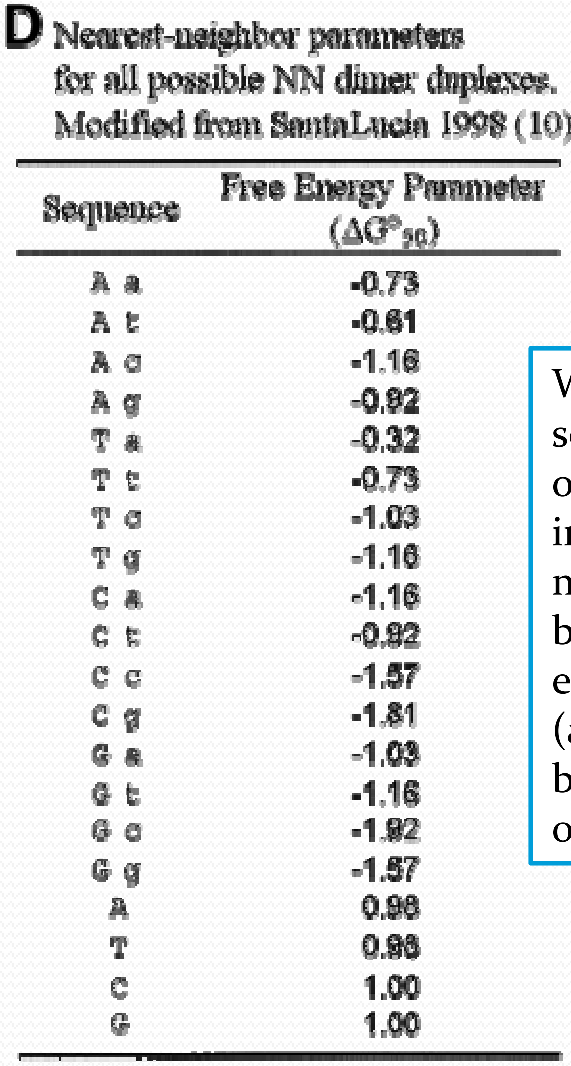
- Advanced options



Dimer check:	This option turns on can avoid primer dimer formation.
Hairpin check:	This option turns on can avoid internal self-complementarity.
5' GC content check:	Check the GC% of 5' to add the ability to recognize the template and enhance the priming specificity.
3' GC content check:	Check the GC% of 3' to avoid mismatch to avoid mismatch.
Covered region:	By entering the start position and stop position, you can get the PCR product containing the segment you need.

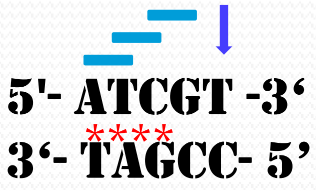
Calculation of the Stability of DNA Duplexes





D Nearest-neighbor parameters for all possible NN dimer duplexes. Modified from SantaLucia 1998 (10).

Sequence	Free Energy Parameter (ΔG°_{50})
A a	-0.73
A t	-0.61
A c	-1.16
A g	-0.92
T a	-0.32
T t	-0.73
T c	-1.03
T g	-1.16
C a	-1.16
C t	-0.82
C c	-1.57
C g	-1.81
G a	-1.03
G t	-1.16
G c	-1.92
G g	-1.57
A	0.98
T	0.98
C	1.00
G	1.00



5'- ATCGT -3'
****
3'- TAGCC- 5'

(-0.61)+(-1.03)+(-1.81)+(0.98)= -2.47

When sequence 1 [5'- atcgt -3'] aligns to sequence 2 [3'- tagcc- 5'], the first base of the first sequence (a) matches to (t) in sequence 2, and follows with three more Watson–Crick pairs. The fifth base mismatches. The NN propagation energy of the continuing base pairs (at), (tc), (cg) and the mismatched base (t) in primer 1 are summed up: (-0.61)+(-1.03)+(-1.81)+(0.98)=-2.47.

Ranking Mechanism

The primer pairs passing through the limitations listed above are sorted by ranking score (*R*):

$$R = 100 - \Delta(T_{\mathrm{m}}) + \Delta G^{\circ}_{\mathrm{forward}}(3' - 5') + \Delta G^{\circ}_{\mathrm{reverse}}(3' - 5') + \text{hairpin score} + \text{dimer score}$$

To avoid the mis-priming amplification, the 5' end of the primer is expected to anneal to target templates more stable than the 3' end

Currently available service (conti)

- **Primer Design Assistant (PDA)**
  - Customized PCR conditions

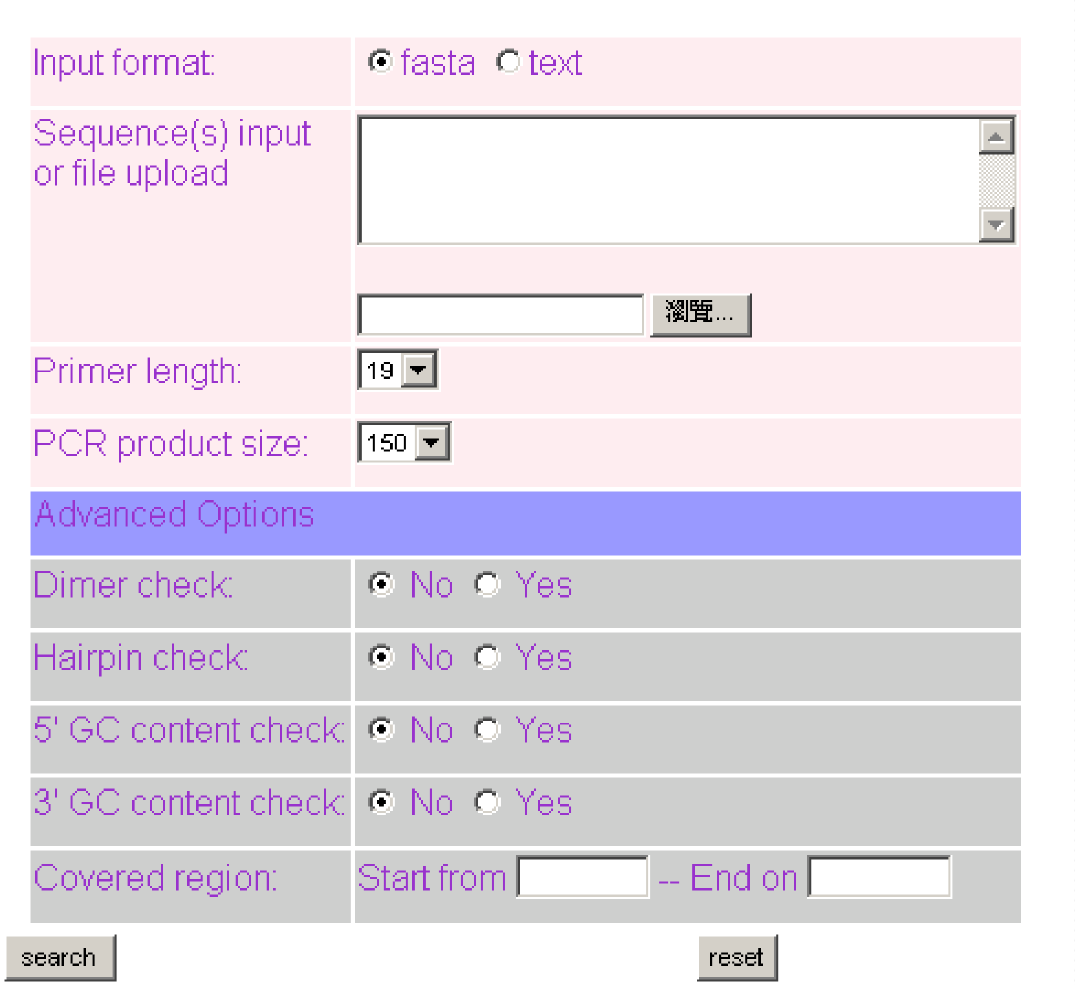
Dimer check

Hairpin check

5’GC content check

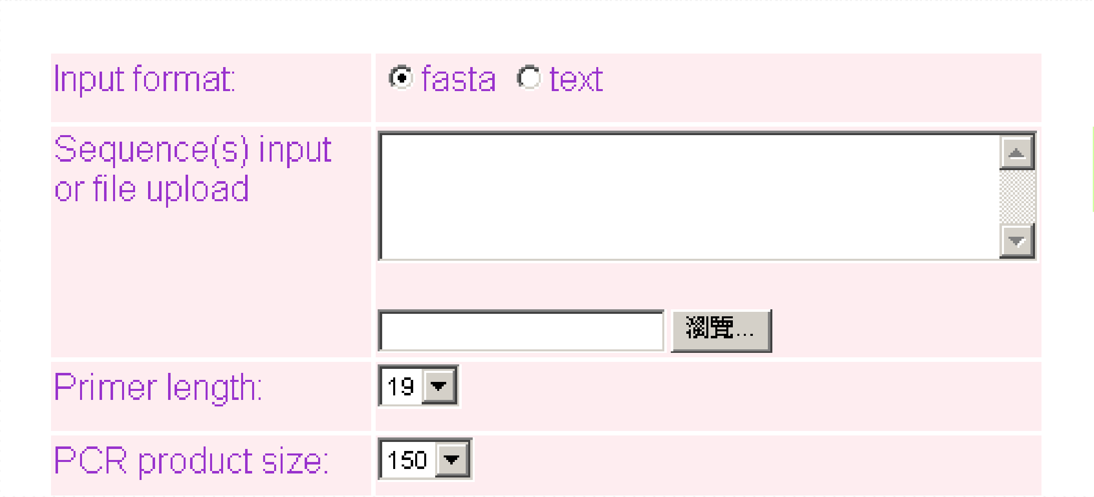
3’GC content check

Covered region



Input format:	● fasta ○ text
Sequence(s) input or file upload	瀏覽...
Primer length:	19
PCR product size:	150
Advanced Options	
Dimer check:	● No ○ Yes
Hairpin check:	● No ○ Yes
5' GC content check:	● No ○ Yes
3' GC content check:	● No ○ Yes
Covered region:	Start from -- End on

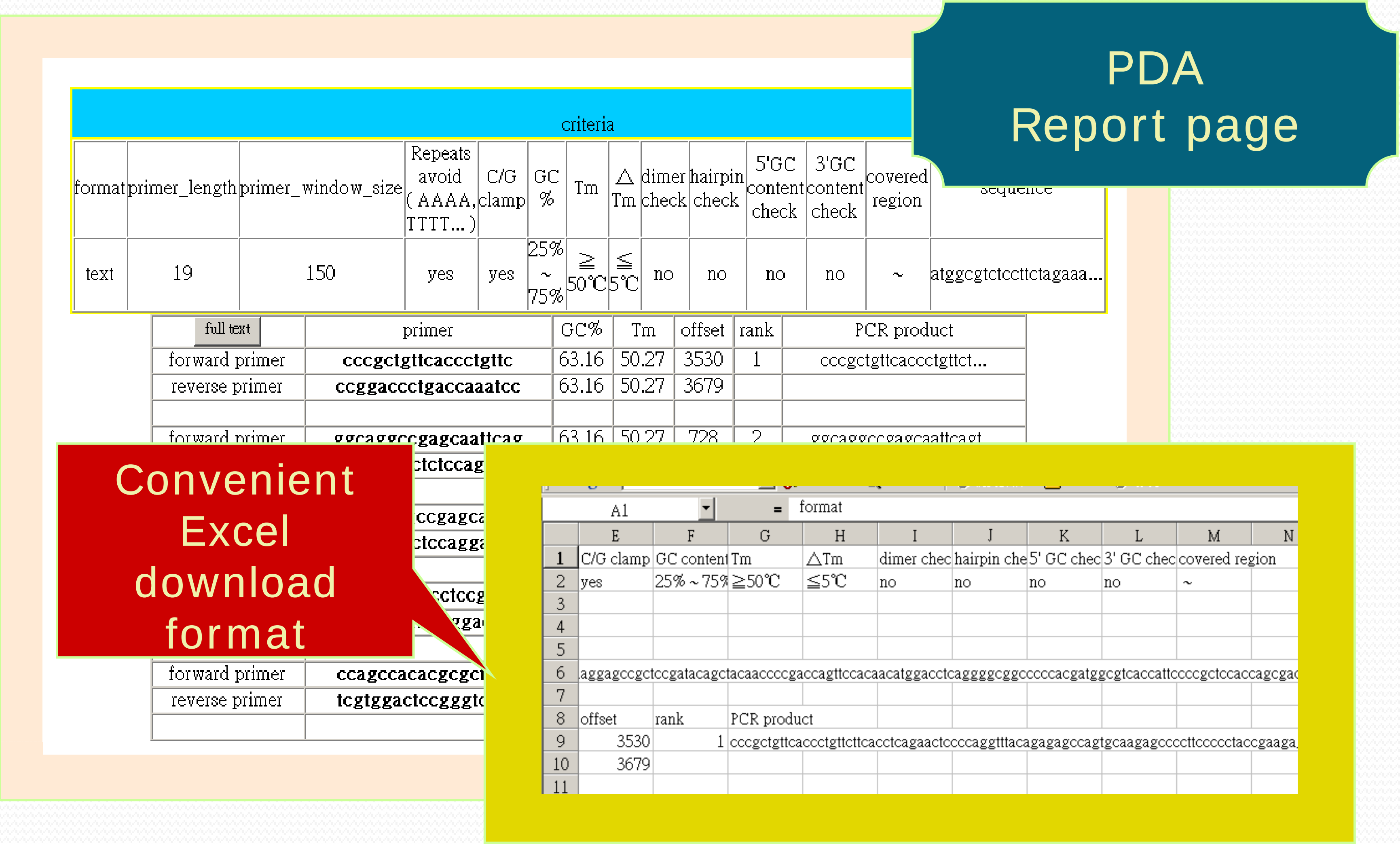
Currently available service (conti)



Input format:	● fasta ○ text
Sequence(s) input or file upload	瀏覽...
Primer length:	19 ▾
PCR product size:	150 ▾

Batch primer design for unified experimental conditions

Currently available service (conti)



criteria														
format	primer_length	primer_window_size	Repeats avoid (AAAA, TTTT...)	C/G clamp	GC %	Tm	△Tm	dimer check	hairpin check	5'GC content check	3'GC content check	covered region	sequence
text	19	150	yes	yes	25% ~ 75%	≧50℃	≦5℃	no	no	no	no	~	atggcgtctccttctagaaa...

full text	primer	GC%	Tm	offset	rank	PCR product
forward primer	**cccgctgttcaccctgttc**	63.16	50.27	3530	1	cccgctgttcaccctgttct...
reverse primer	**ccggaccctgaccaaatcc**	63.16	50.27	3679		
forward primer	**ggcaggccgagcaattcag**	63.16	50.27	728	2	ggcaggccgagcaattcagt
	ctctccag					
	ccgagca					
	ctccagga					
	cctccg					
	gga					
forward primer	**ccagccacacgcgc**					
reverse primer	**tcgtggactccgggtc**					

A1 = format

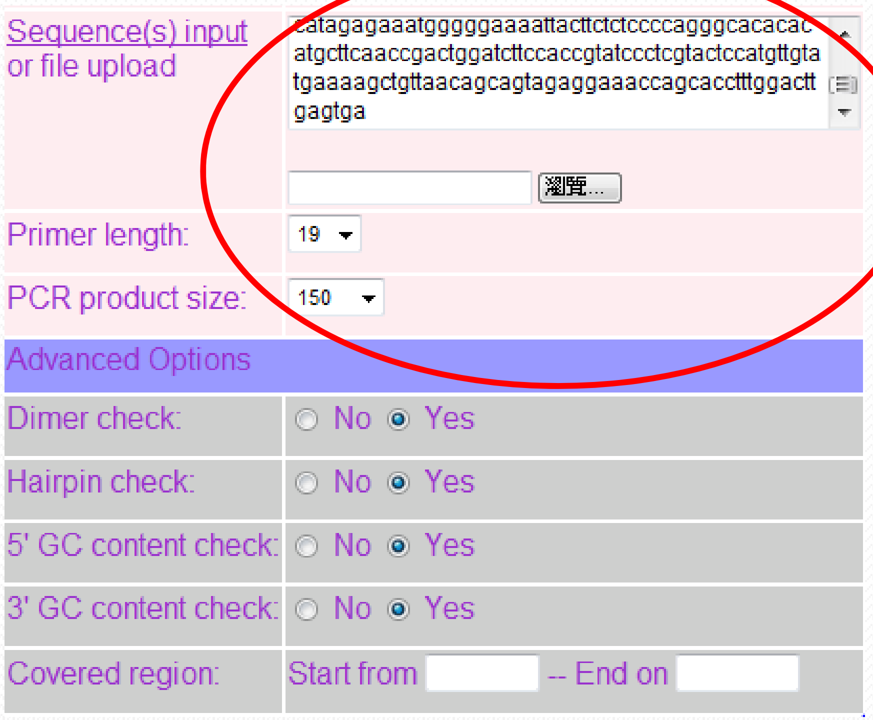
	E	F	G	H	I	J	K	L	M	N
1	C/G clamp	GC content	Tm	△Tm	dimer chec	hairpin che	5' GC chec	3' GC chec	covered region	
2	yes	25% ~ 75%	≧50℃	≦5℃	no	no	no	no	~	
3										
4										
5										
6	aggagccgctccgatacagctacaaccccgaccagttccacaacatggacctcaggggcggcccccacgatggcgtcaccattccccgctccaccagcgac									
7										
8	offset	rank	PCR product							
9	3530	1	cccgctgttcaccctgttcttcacctcagaactccccaggtttacagagagccagtgcaagagccccttccccctaccgaaga							
10	3679									
11										

PDA
Report page

Convenient
Excel
download
format

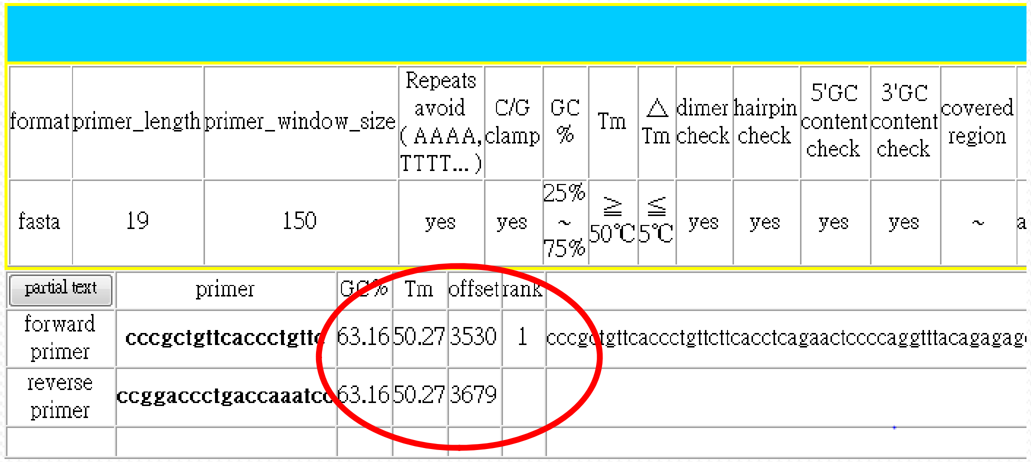
Primer set for Nested PCR in PDA

Input target sequence



Sequence(s) input or file upload	catagagaaatggggaaaattacttctctccccagggcacacac atgcttcaaccgactggatcttccaccgtatccctcgtactccatgttgta tgaaaagctgttaacagcagtagaggaaaccagcacctttggactt gagtga 瀏覽...
Primer length:	19
PCR product size:	150
Advanced Options	
Dimer check:	○ No ● Yes
Hairpin check:	○ No ● Yes
5' GC content check:	○ No ● Yes
3' GC content check:	○ No ● Yes
Covered region:	Start from -- End on

Get the primer set and their location

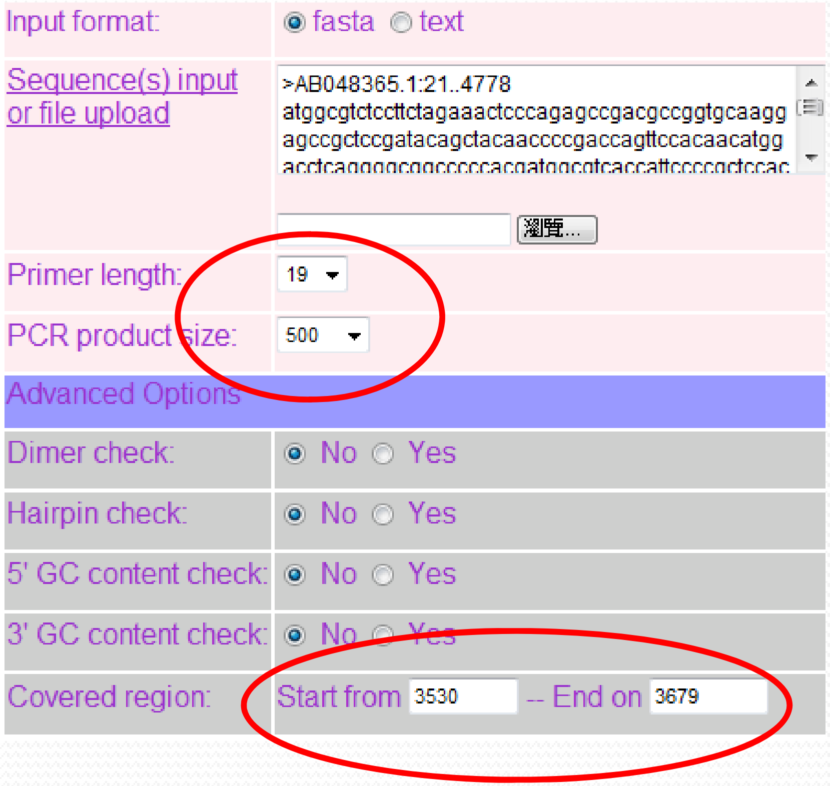


format	primer_length	primer_window_size	Repeats avoid (AAAA, TTTT...)	C/G clamp	GC %	Tm	△ Tm	dimer check	hairpin check	5'GC content check	3'GC content check	covered region	
fasta	19	150	yes	yes	25% ~ 75%	≧ 50℃	≦ 5℃	yes	yes	yes	yes	~	a

partial text	primer	GC%	Tm	offset	rank	
forward primer	**cccgctgttcaccctgttc**	63.16	50.27	3530	1	cccgctgttcaccctgttcttcacctcagaactccccaggtttacagagag
reverse primer	**ccggaccctgaccaaatcc**	63.16	50.27	3679		

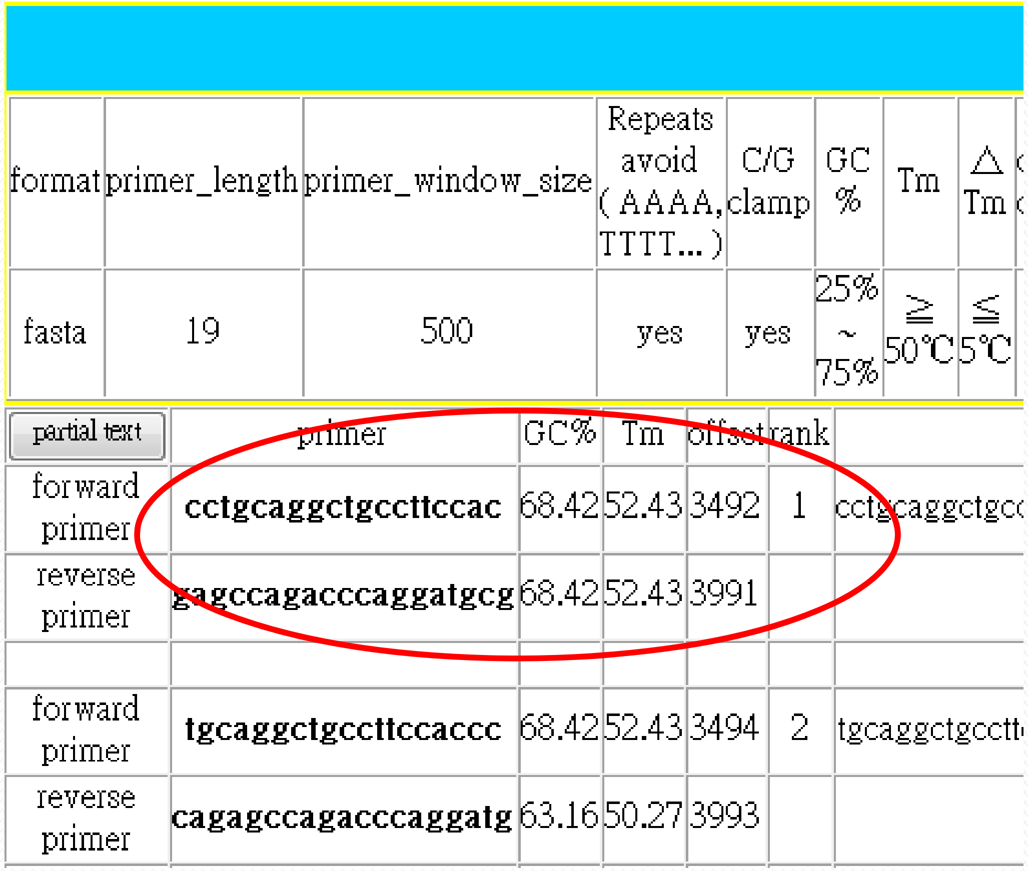
Primer set for Nested PCR in PDA

Modify the size of Product and fill the location of 1st PCR product



Input format:	● fasta ○ text
Sequence(s) input or file upload	>AB048365.1:21..4778 atggcgtctccttctagaaactcccagagccgacgccggtgcaagg agccgctccgatacagctacaaccccgaccagttccacaacatgg acctcaggggcggccccacgatggcgtcaccattccccgctccac 瀏覽...
Primer length:	19
PCR product size:	500
Advanced Options	
Dimer check:	● No ○ Yes
Hairpin check:	● No ○ Yes
5' GC content check:	● No ○ Yes
3' GC content check:	● No ○ Yes
Covered region:	Start from 3530 -- End on 3679

Get the primer sets for nested PCR with 500 bps product

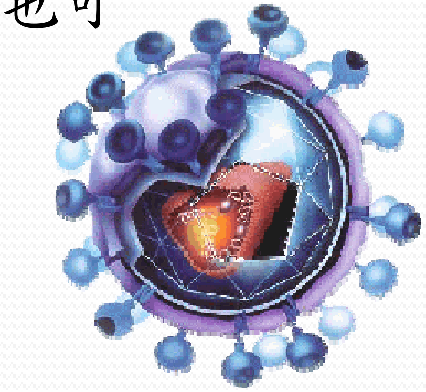


format	primer_length	primer_window_size	Repeats avoid (AAAA, TTTT...)	C/G clamp	GC %	Tm	△Tm
fasta	19	500	yes	yes	25% ~ 75%	≧ 50℃	≦ 5℃

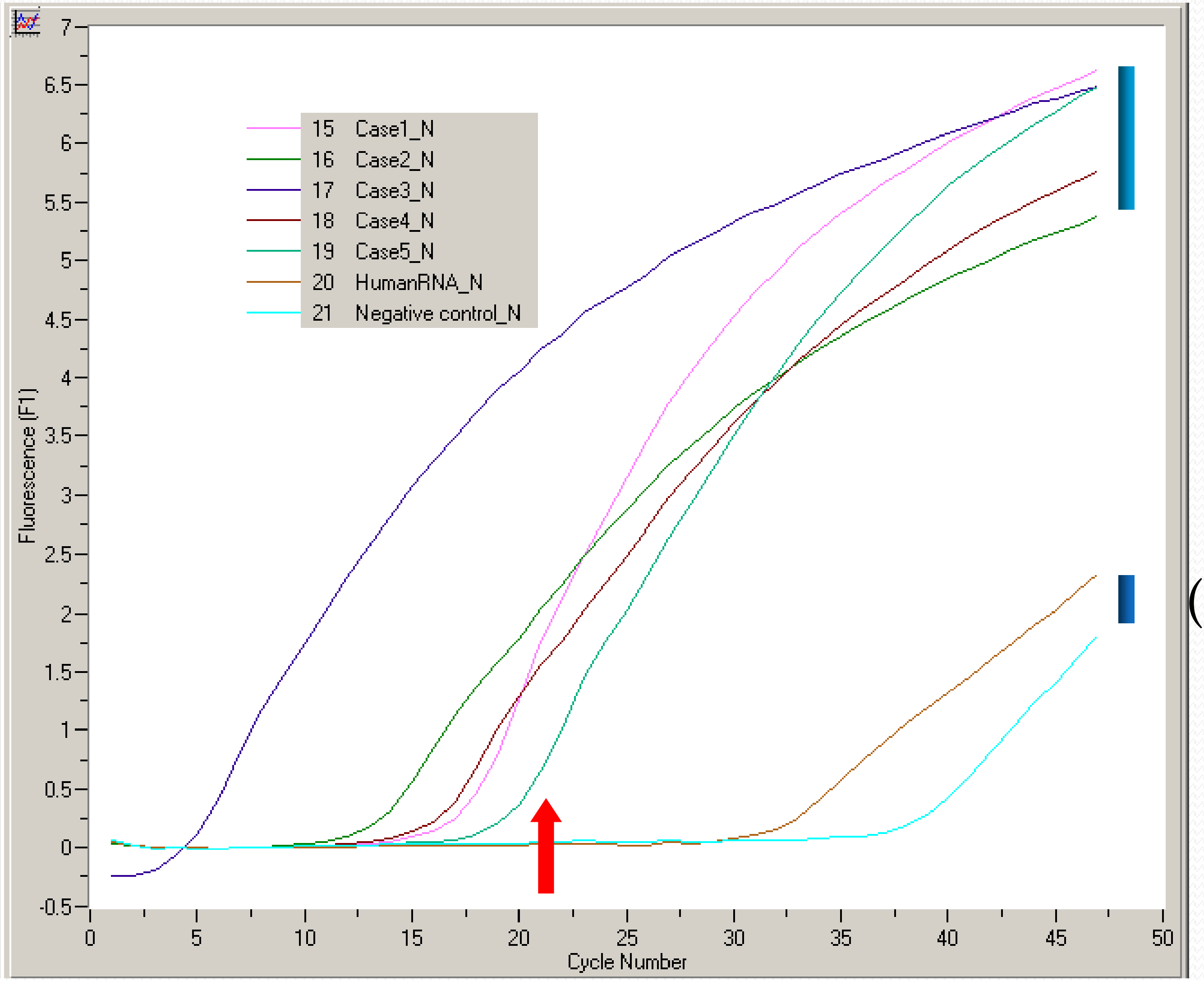
partial text	primer	GC%	Tm	offset	rank	
forward primer	cctgcaggctgccttccac	68.42	52.43	3492	1	cctgcaggctgcc
reverse primer	gagccagacccaggatgcg	68.42	52.43	3991		
forward primer	tgcaggctgccttccaccc	68.42	52.43	3494	2	tgcaggctgcctt
reverse primer	cagagccagacccaggatg	63.16	50.27	3993		

Use PDA to Develop PCR kits for SARS Detection

- 由於2003年當時通用的SARS-CoV檢驗方法靈敏度有限，使得病毒量較低或是因採樣方法不佳的檢體無法被檢測到，在防疫的前提下，本核心便與國家衛生研究院基因醫學研究組協同疾病管制局(CDC, Taiwan)發展出高靈敏度之檢測方法。
- 檢測方法中所需要的核酸引子都透過PDA來進行設計，避免引子本身dimer 及 hairpins 的形成，加速了檢測方法的建立。
- 此方法為結合1st run RT-PCR + 2nd run Q-PCR，可於1.5小時內檢測出結果，經實驗證明縱使病毒量低於10隻，也可以透過這一套方法檢測出來。



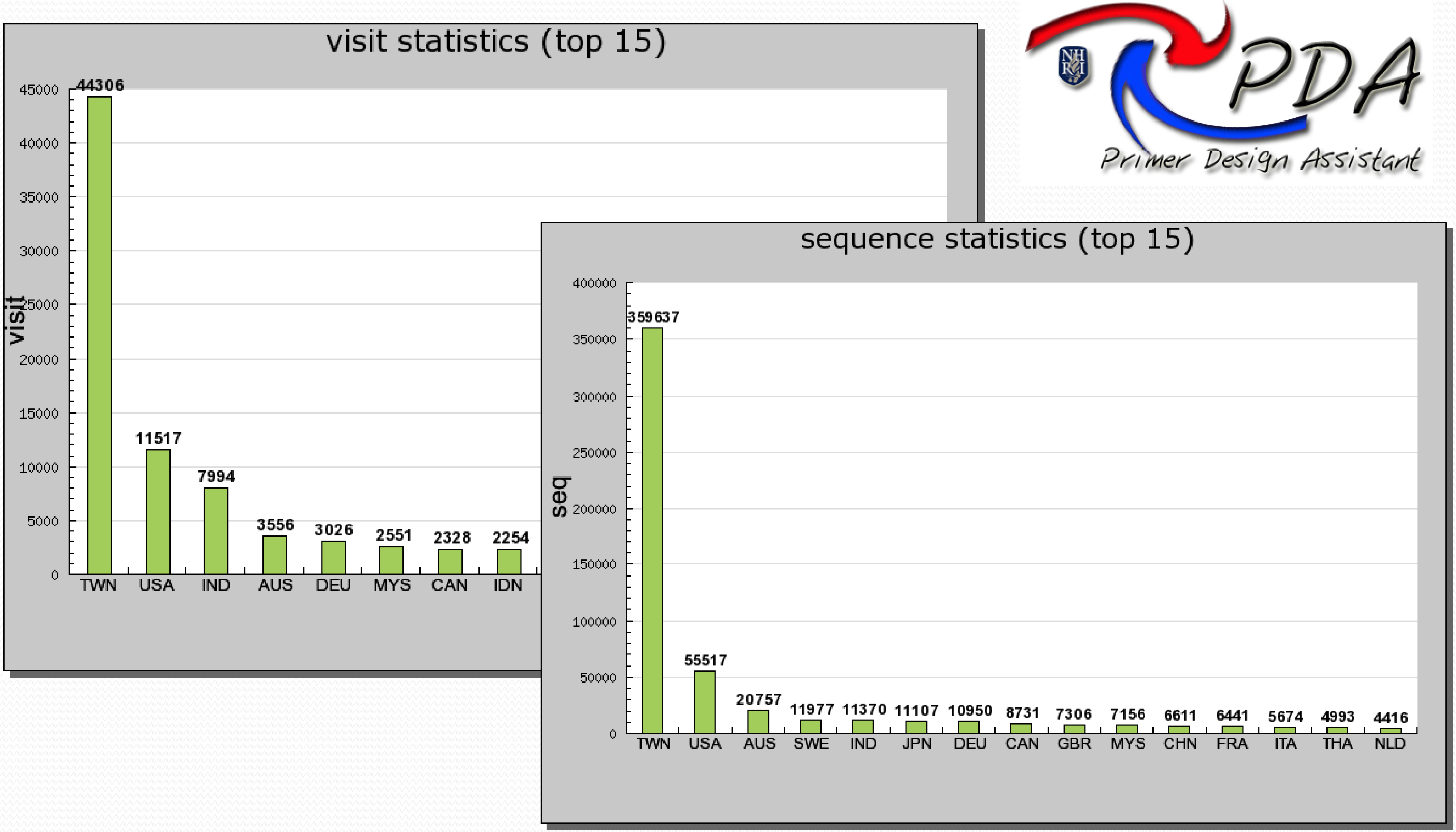
Result of Real-time PCR for SARS-CoV Detection



Cases

(-) Control

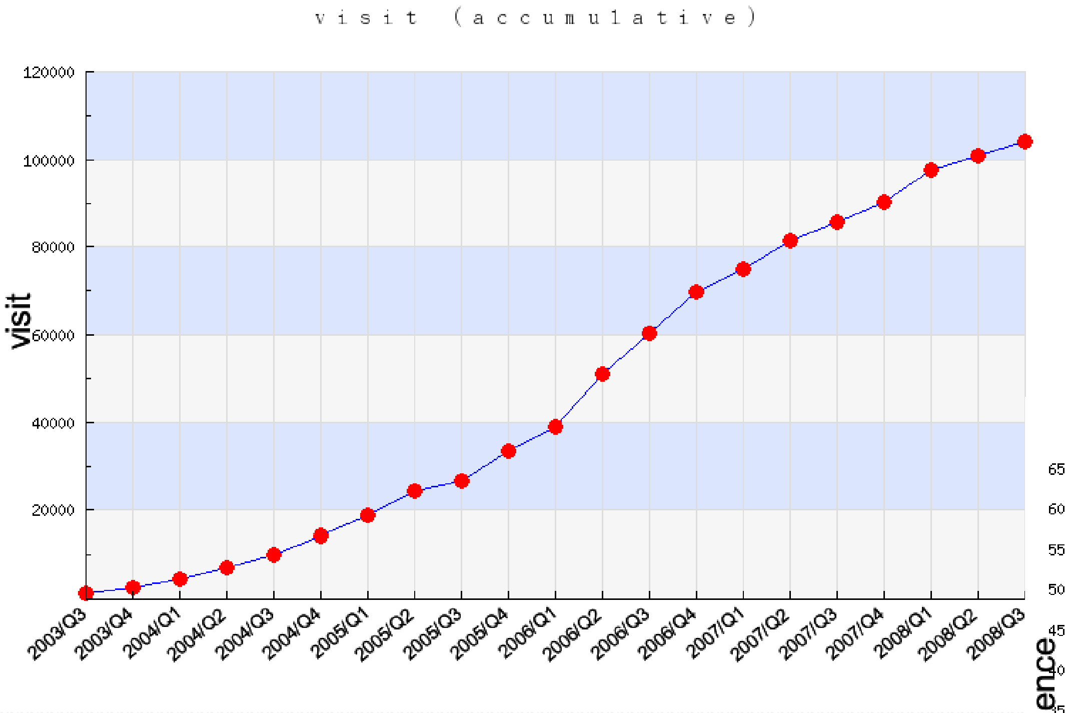
Primer Design Assistant (PDA)

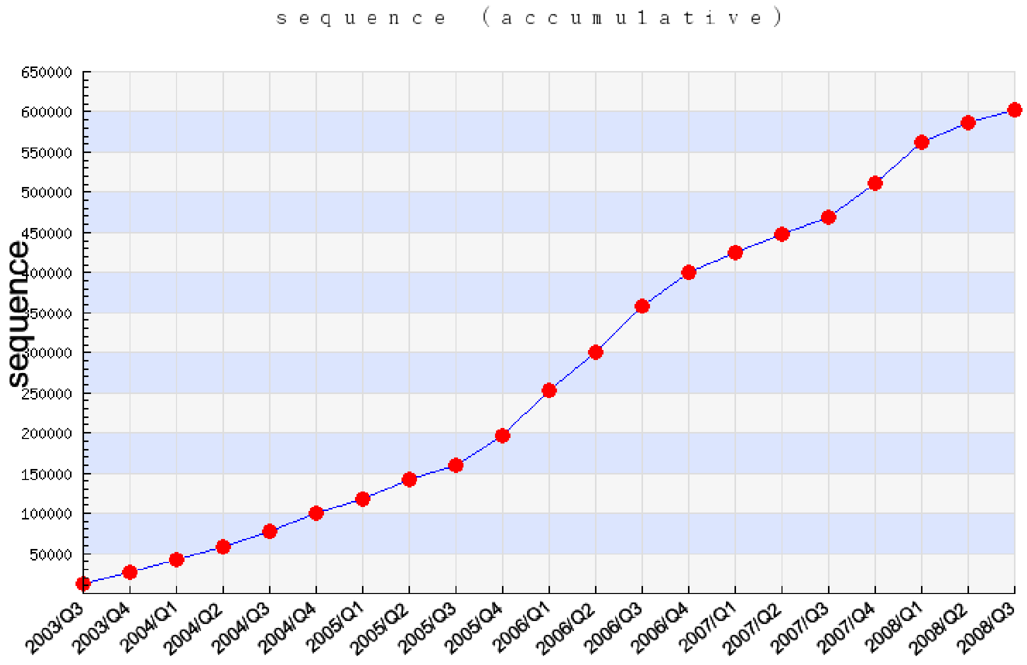


From July 2003 to Aug 2008

Nucleic Acids Res. 2003, 31: 3751-3754

Primer Design Assistant (PDA)

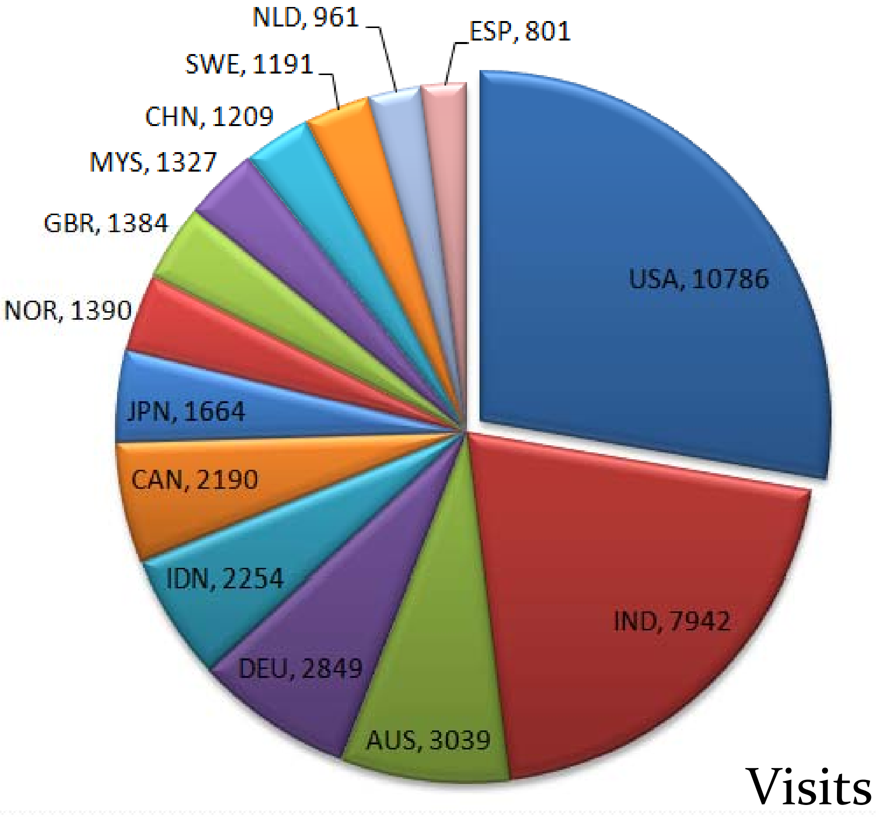


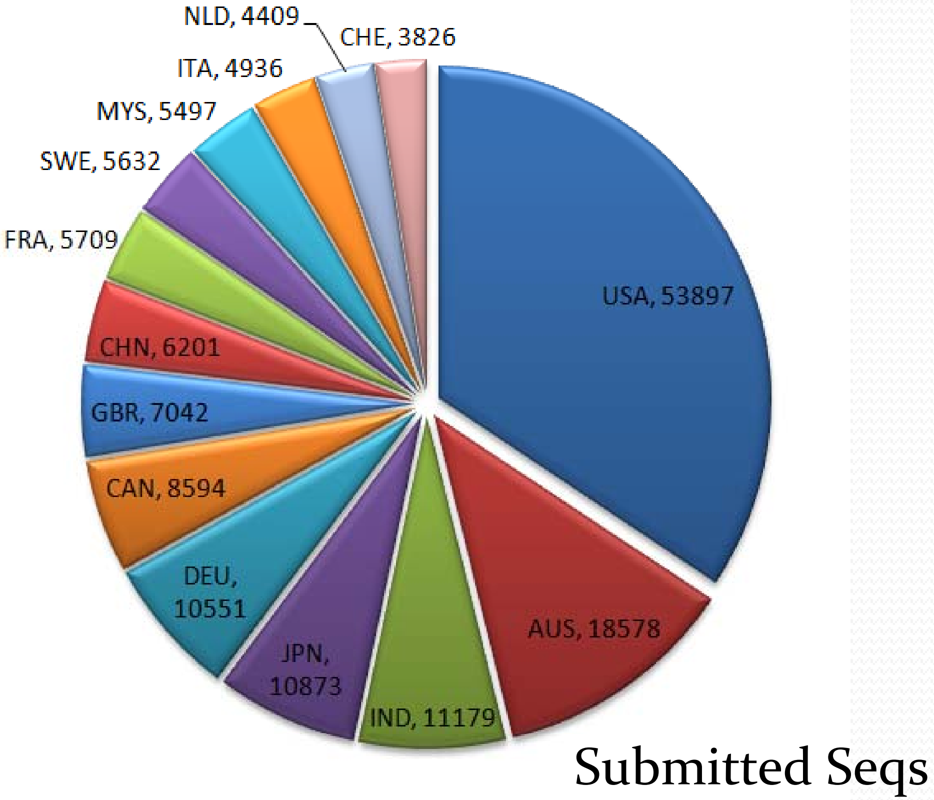


From July 2003 to Aug 2008
Over 100,000 Visits and 600,000 submitted Seqs

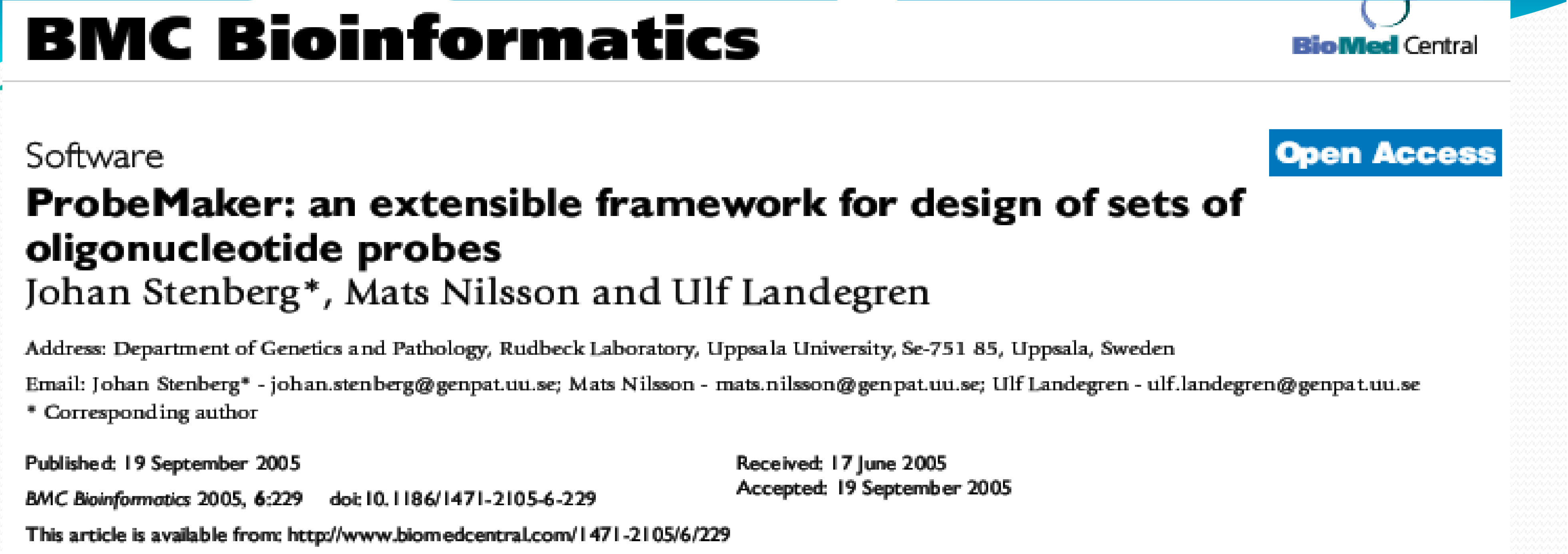
Nucleic Acids Res. 2003, 31: 3751-3754

Usage of PDA Worldwide





Sequences submitted to PDA from overseas, accumulative,
Jul. 2003 to Apr. 2008; with 334,426 sequences submitted from Taiwan



BMC Bioinformatics

BioMed Central

Software

Open Access

ProbeMaker: an extensible framework for design of sets of oligonucleotide probes

Johan Stenberg*, Mats Nilsson and Ulf Landegren

Address: Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Se-751 85, Uppsala, Sweden

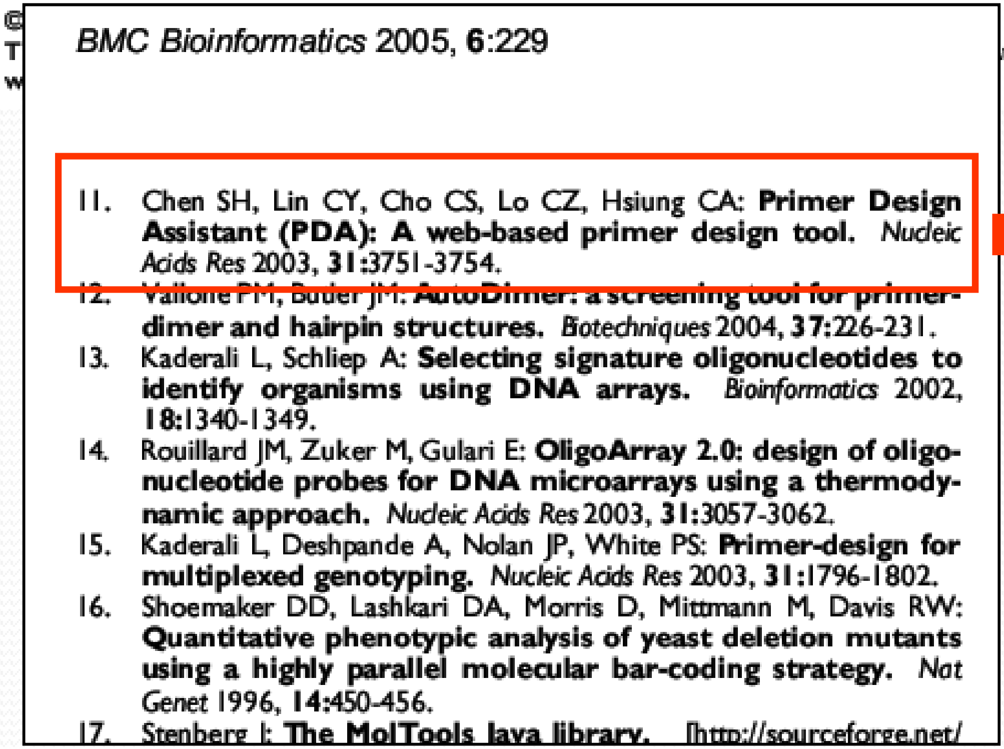
Email: Johan Stenberg* - johan.stenberg@genpat.uu.se; Mats Nilsson - mats.nilsson@genpat.uu.se; Ulf Landegren - ulf.landegren@genpat.uu.se
* Corresponding author

Published: 19 September 2005

BMC Bioinformatics 2005, 6:229 doi:10.1186/1471-2105-6-229

Received: 17 June 2005
Accepted: 19 September 2005

This article is available from: http://www.biomedcentral.com/1471-2105/6/229



BMC Bioinformatics 2005, **6**:229

11. Chen SH, Lin CY, Cho CS, Lo CZ, Hsiung CA: **Primer Design Assistant (PDA): A web-based primer design tool.** *Nucleic Acids Res* 2003, **31**:3751-3754.

12. Vallone PM, Butler JM: **AutoDimer: a screening tool for primer-dimer and hairpin structures.** *Biotechniques* 2004, **37**:226-231.

13. Kaderali L, Schliep A: **Selecting signature oligonucleotides to identify organisms using DNA arrays.** *Bioinformatics* 2002, **18**:1340-1349.

14. Rouillard JM, Zuker M, Gulari E: **OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach.** *Nucleic Acids Res* 2003, **31**:3057-3062.

15. Kaderali L, Deshpande A, Nolan JP, White PS: **Primer-design for multiplexed genotyping.** *Nucleic Acids Res* 2003, **31**:1796-1802.

16. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW: **Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy.** *Nat Genet* 1996, **14**:450-456.

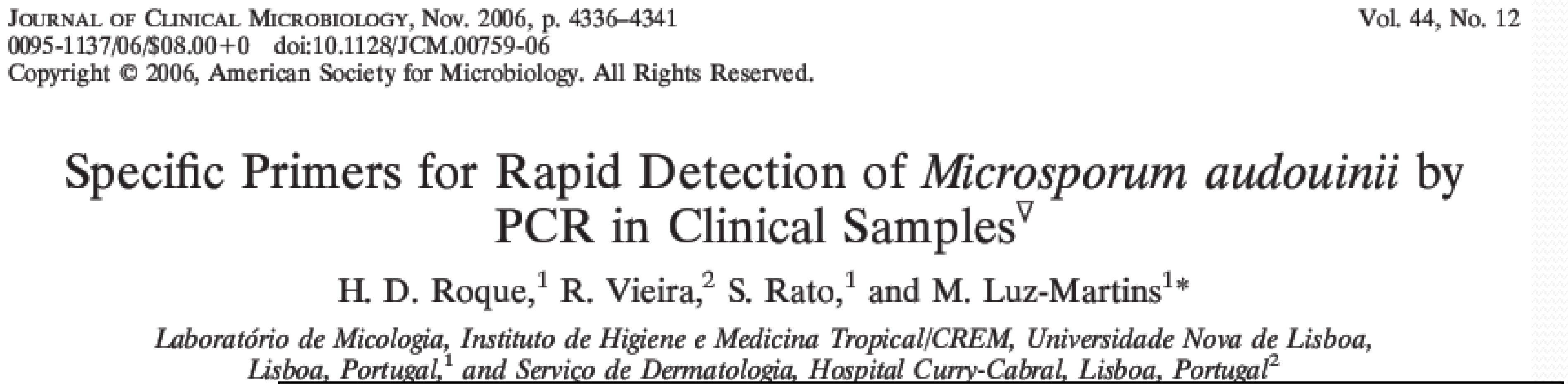
17. Stenberg J: **The MolTools Java library.** [http://sourceforge.net/

Chen SH, Lin CY, Cho CS, Lo CZ, Hsiung CA: Primer Design Assistant (PDA): A web-based primer design tool. Nucleic Acids Res 2003, 31:3751-3754.

Input format: ● fasta ○ text

Primer length: 19

PCR product size: 150

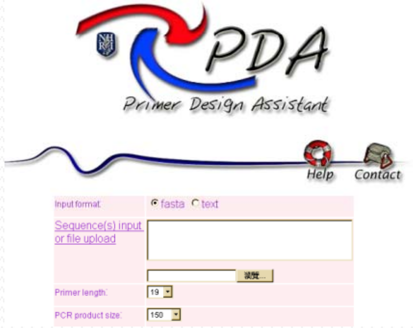


Journal of Clinical Microbiology, Nov. 2006, p. 4336–4341 Vol. 44, No. 12
0095-1137/06/$08.00+0 doi:10.1128/JCM.00759-06

Specific Primers for Rapid Detection of *Microsporum audouinii* by PCR in Clinical Samples[▽]

H. D. Roque,[1] R. Vieira,[2] S. Rato,[1] and M. Luz-Martins[1]*

Laboratório de Micologia, Instituto de Higiene e Medicina Tropical/CREM, Universidade Nova de Lisboa, Lisboa, Portugal,[1] and Serviço de Dermatologia, Hospital Curry-Cabral, Lisboa, Portugal[2]

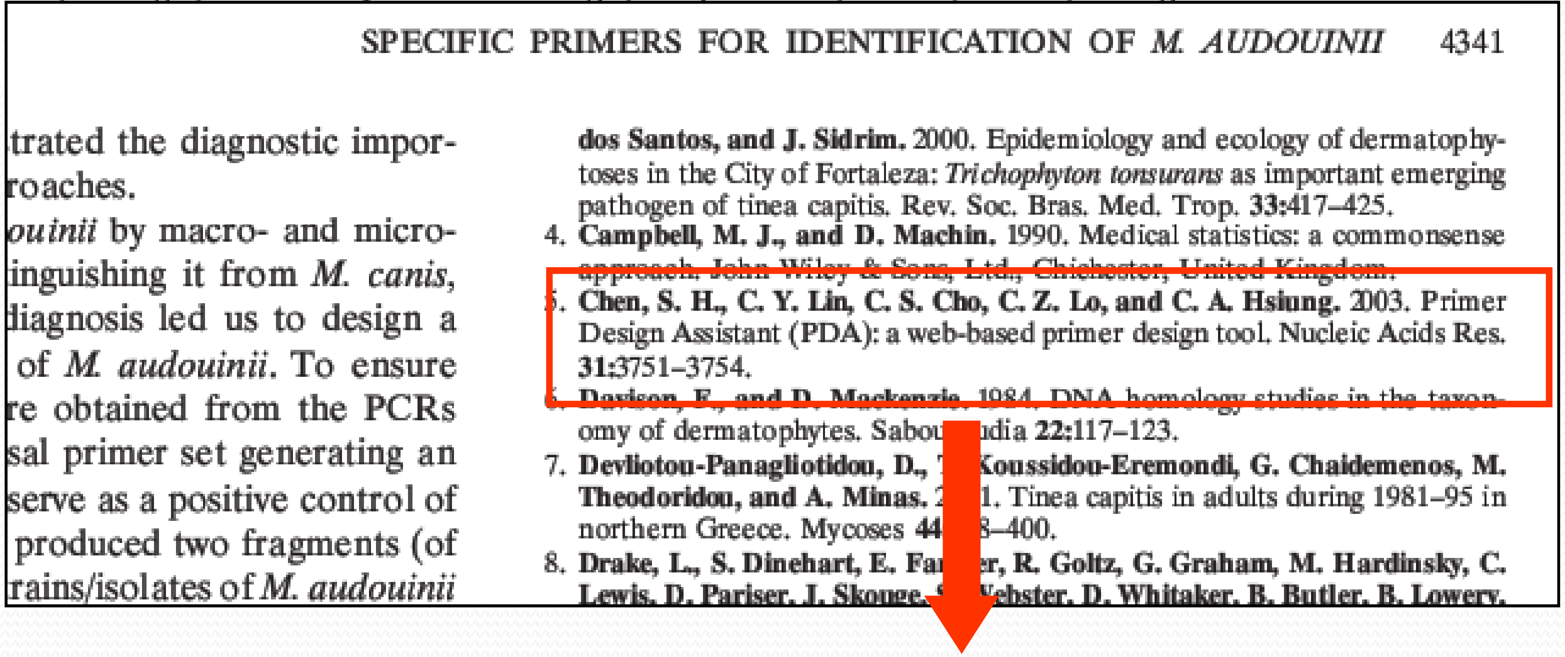


Input format: ● fasta ○ text

Sequence(s) input or file upload

Primer length: 19

PCR product size: 150



SPECIFIC PRIMERS FOR IDENTIFICATION OF *M. AUDOUINII* 4341

trated the diagnostic impor-
roaches.
ouinii by macro- and micro-
inguishing it from *M. canis*,
diagnosis led us to design a
of *M. audouinii*. To ensure
re obtained from the PCRs
sal primer set generating an
serve as a positive control of
produced two fragments (of
rains/isolates of *M. audouinii*

dos Santos, and J. Sidrim. 2000. Epidemiology and ecology of dermatophytoses in the City of Fortaleza: *Trichophyton tonsurans* as important emerging pathogen of tinea capitis. Rev. Soc. Bras. Med. Trop. **33:**417–425.

4. **Campbell, M. J., and D. Machin.** 1990. Medical statistics: a commonsense approach. John Wiley & Sons, Ltd., Chichester, United Kingdom.

5. **Chen, S. H., C. Y. Lin, C. S. Cho, C. Z. Lo, and C. A. Hsiung.** 2003. Primer Design Assistant (PDA): a web-based primer design tool. Nucleic Acids Res. **31:**3751–3754.

6. **Davison, F., and D. Mackenzie.** 1984. DNA homology studies in the taxonomy of dermatophytes. Sabou[illegible]dia **22:**117–123.

7. **Devliotou-Panagliotidou, D., T[illegible] Koussidou-Eremondi, G. Chaidemenos, M. Theodoridou, and A. Minas.** 2[illegible]1. Tinea capitis in adults during 1981–95 in northern Greece. Mycoses **44**[illegible]8–400.

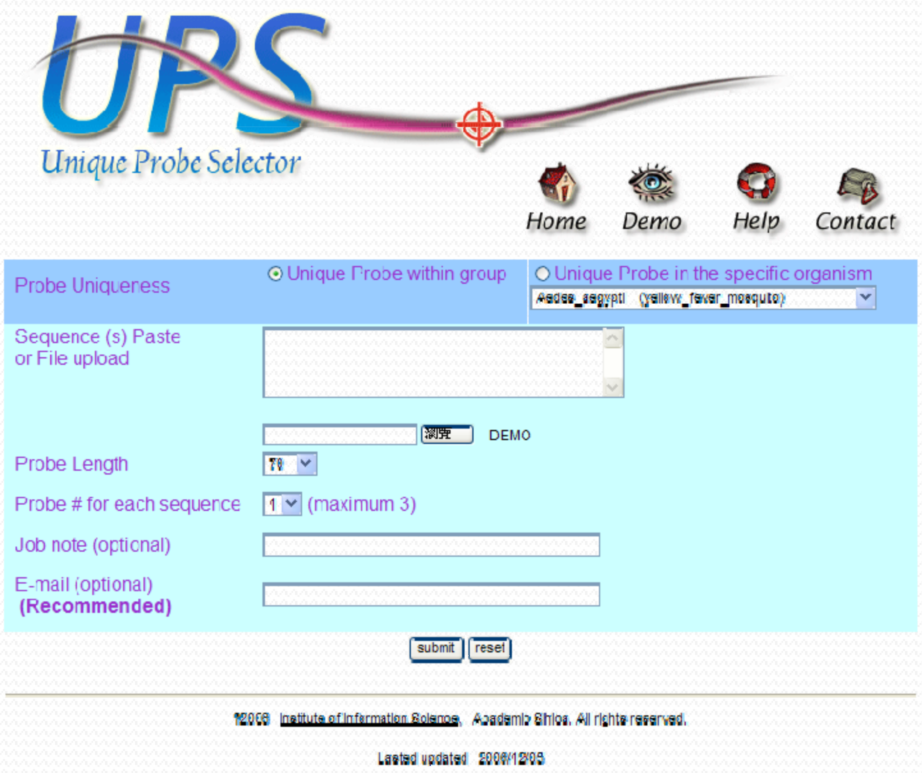
8. **Drake, L., S. Dinehart, E. Far[illegible]er, R. Goltz, G. Graham, M. Hardinsky, C. Lewis, D. Pariser, J. Skouge, [illegible] Webster, D. Whitaker, B. Butler, B. Lowery,**

Chen, S. H., C. Y. Lin, C. S. Cho, C. Z. Lo, and C. A. Hsiung. 2003. Primer Design Assistant (PDA): a web-based primer design tool. Nucleic Acids Res. 31:3751–3754

Genomics Studies For High Throughput Research : UPS

BMC Bioinformatics

BMC Bioinformatics 2007, **9**(Suppl 1):S8



Probe Uniqueness	● Unique Probe within group	○ Unique Probe in the specific organism Aedes_aegypti (yellow_fever_mosquito)
Sequence (s) Paste or File upload		
	瀏覽 DEMO	
Probe Length	70	
Probe # for each sequence	1 (maximum 3)	
Job note (optional)		
E-mail (optional) **(Recommended)**		

submit reset

©2006 Institute of Information Science, Academia Sinica. All rights reserved.

Lasted updated 2006/12/05

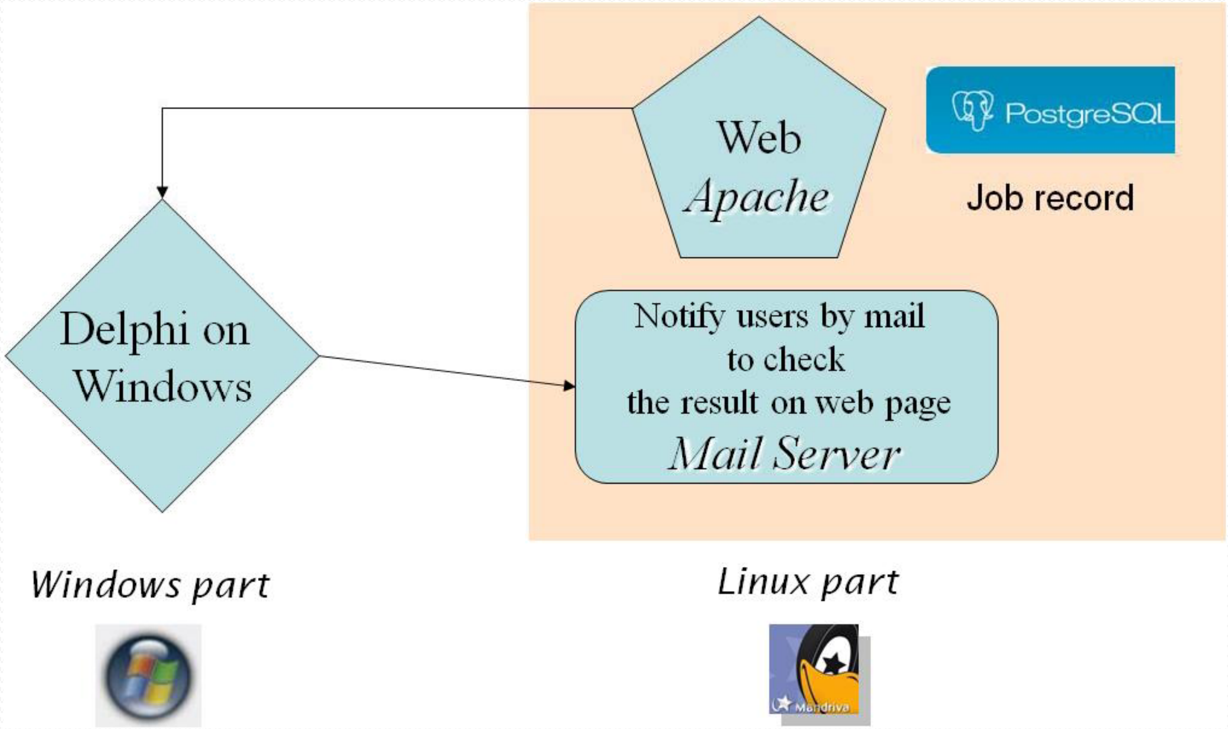
- **Unique Probe Selector (UPS)**
- ***Probe design for hybridization in low and high throughput experiments***
- **http://array.iis.sinica.edu.tw/ups/**
- **BMC Bioinformatics 2008**

Brief on Unique Probe Selector (UPS)

- Although most of these tools aimed on designing probes for microarray, only few of them take the genetic background noise in hybridization reaction into account. **A web tool** for customized probe design regarding to the discriminating power of probes is sparse.
- Here we present a web tool, the Unique Probe Selector (UPS), for selecting unique oligo-nucleotide probe. The algorithms applied here include **thermodynamic theory**, **GC content**, **GC clamps**, **secondary structure of probes** and **some other empirical preferences of wet-lab researchers**. **Low-complexity** regions are filtered out to maintain probe specificity.
- The UPS evaluates probe-to-target hybridization under a user-defined condition *in silico* **to ensure high-performance hybridization** and **minimizes the possibility of non-specific reactions**.
- UPS has been applied to design human arrays for gene expression studies and to develop several small arrays of gene families that were inferred as molecular signatures of cancer typing/staging or pathogen signatures.

Infrastructure of UPS

- Under the consideration of efficiency and performance, we adopted the LAMP structure (Mandrake 2007, the operating system , Apache (webserver), PostgreSQL (database) and PHP), to provide the web access, files upload/ download, mail notification and data storage.
- All the calculations related with unique probe design was performed on a window-based machine in Delphi code.



Basic Criteria for Probe Selection

① Probe length: from 30 ~ 120 bps

② Melting Temperature

- ✓ The probe annealing temperature (Ta) is determined by melting temperature (Tm). Probe Tm depends on several physiochemical factors and is calculated in the following equation based on Nearest Neighbor model

$$\Delta Tm = \Delta H / (10.8 + \Delta S + R \times \ln (C / 4)) - 273.15 + 16.6(\log_{10}[Salt])$$

③ Sequence complexity

- ✓ We exclude any five or more continual nucleotides (AAAAA, TTTTT, CCCCC, or GGGGG). Continuous di-nucleotide/ tri-nucleotide repeats, such as 'ATATAT' and 'ATGATGATG', are also avoided.

Basic Criteria for Probe Selection

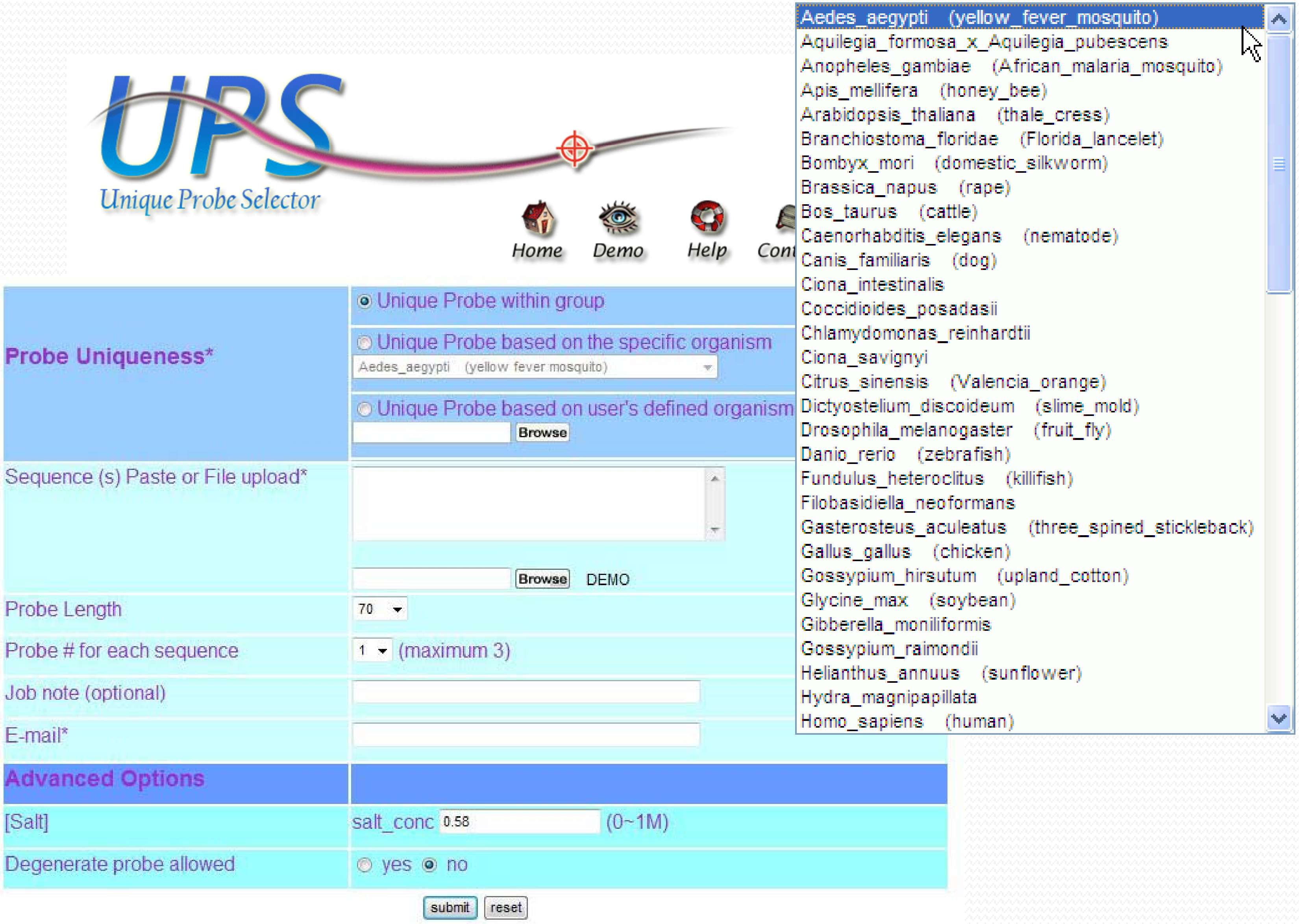
④ Computation of secondary structure formation

- ✓ We use a perl program UNAFold.pl integrated into UPS to calculate ΔG

⑤ Continuous stretch and identity between probe and no-target template

- ✓ Here we used Li *et al's* (NAR 2003) experimentally established criteria to exclude unsuitable oligonucleotides: **identity of ≧ 85%, continuous stretch of ≧17 and free energy <-35 kcal/mol** (it will depend on the length of probe) between probe and non-target templates.

Demonstration of UPS 2.0

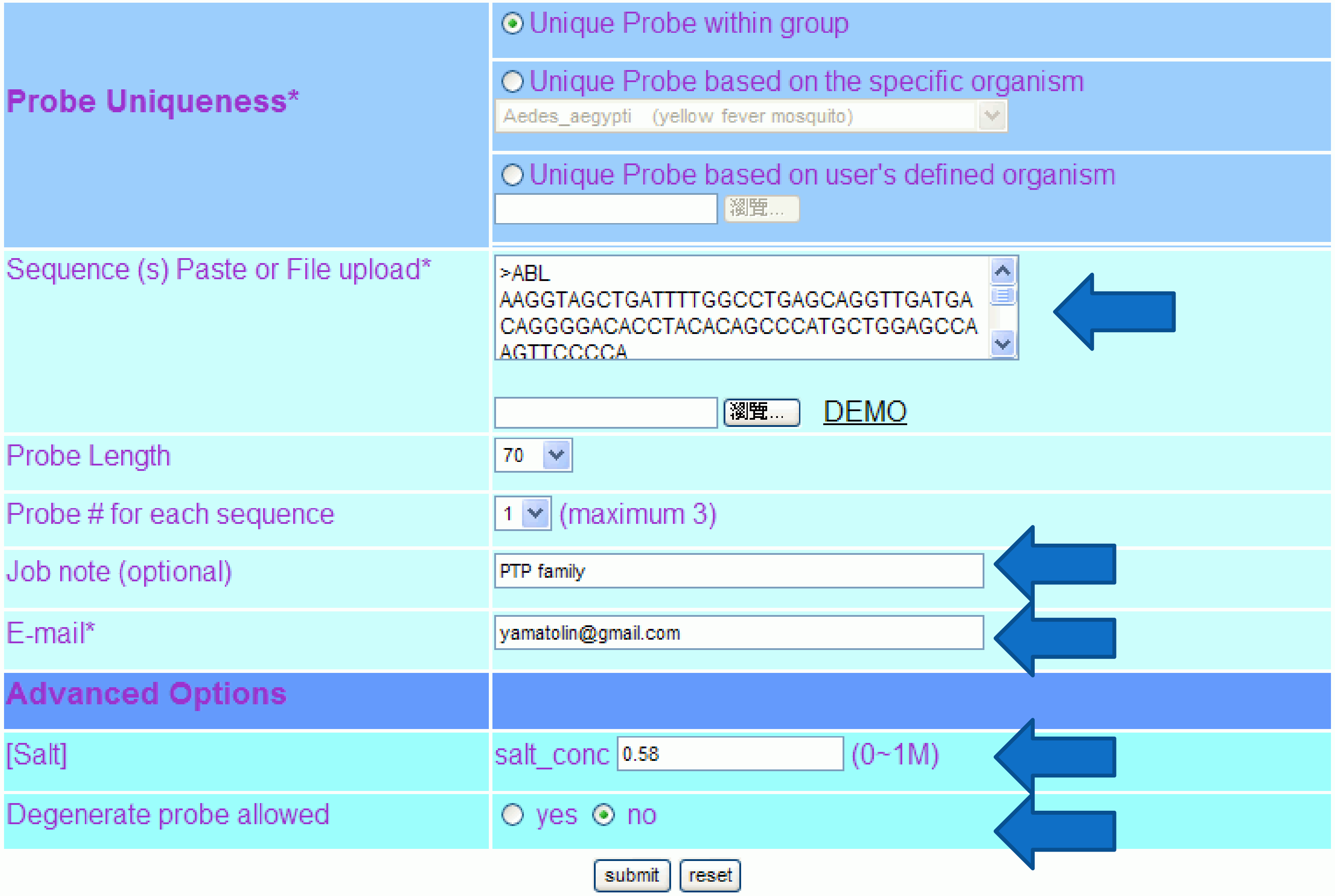


Probe Uniqueness*	● Unique Probe within group ○ Unique Probe based on the specific organism Aedes_aegypti (yellow fever mosquito) ○ Unique Probe based on user's defined organism Browse
Sequence (s) Paste or File upload*	Browse DEMO
Probe Length	70
Probe # for each sequence	1 (maximum 3)
Job note (optional)	
E-mail*	
Advanced Options	
[Salt]	salt_conc 0.58 (0~1M)
Degenerate probe allowed	○ yes ● no

submit reset

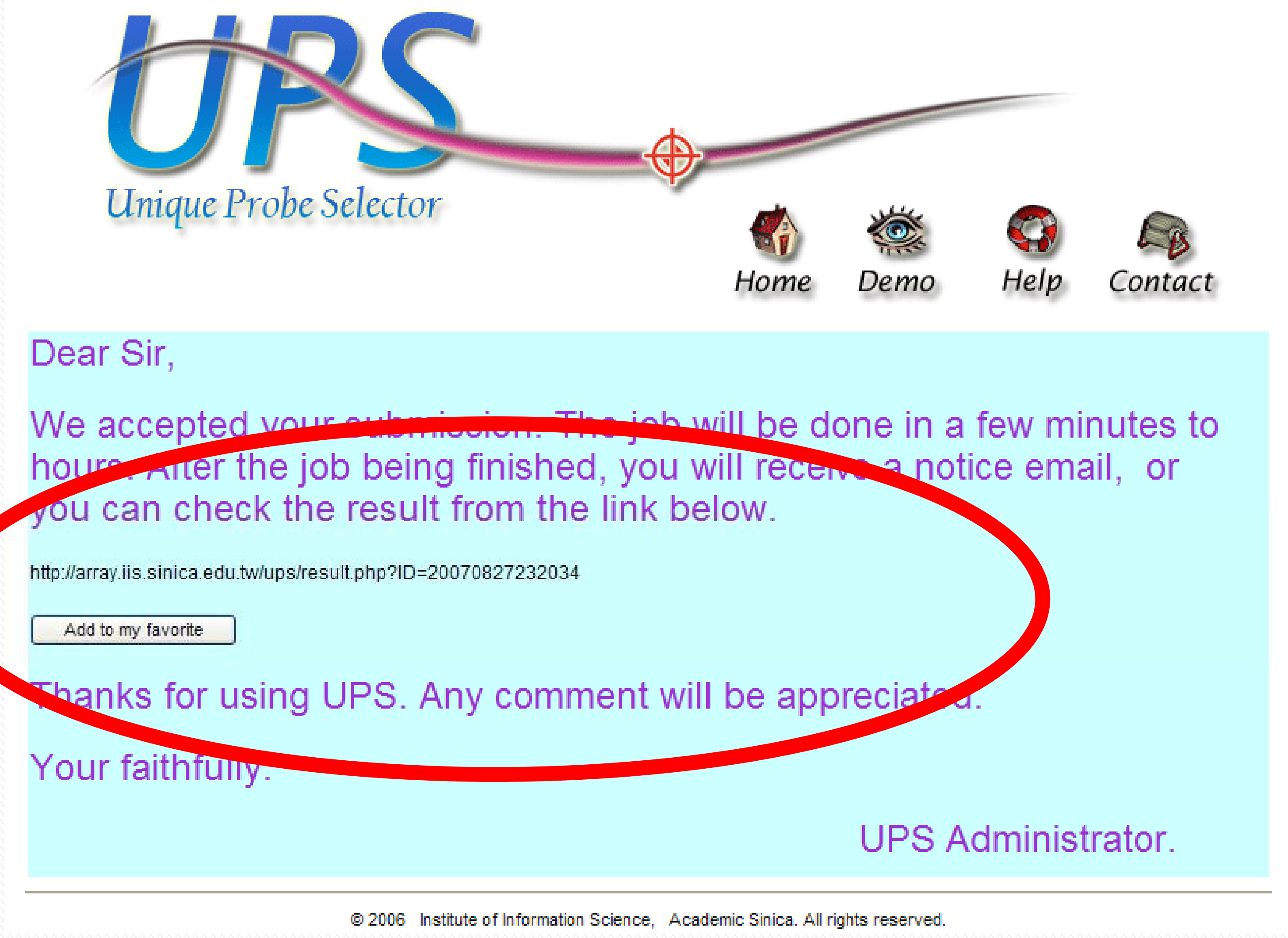
Aedes_aegypti (yellow_fever_mosquito)
Aquilegia_formosa_x_Aquilegia_pubescens
Anopheles_gambiae (African_malaria_mosquito)
Apis_mellifera (honey_bee)
Arabidopsis_thaliana (thale_cress)
Branchiostoma_floridae (Florida_lancelet)
Bombyx_mori (domestic_silkworm)
Brassica_napus (rape)
Bos_taurus (cattle)
Caenorhabditis_elegans (nematode)
Canis_familiaris (dog)
Ciona_intestinalis
Coccidioides_posadasii
Chlamydomonas_reinhardtii
Ciona_savignyi
Citrus_sinensis (Valencia_orange)
Dictyostelium_discoideum (slime_mold)
Drosophila_melanogaster (fruit_fly)
Danio_rerio (zebrafish)
Fundulus_heteroclitus (killifish)
Filobasidiella_neoformans
Gasterosteus_aculeatus (three_spined_stickleback)
Gallus_gallus (chicken)
Gossypium_hirsutum (upland_cotton)
Glycine_max (soybean)
Gibberella_moniliformis
Gossypium_raimondii
Helianthus_annuus (sunflower)
Hydra_magnipapillata
Homo_sapiens (human)

Demonstration of UPS 2.0



Probe Uniqueness*	● Unique Probe within group
	○ Unique Probe based on the specific organism Aedes_aegypti (yellow fever mosquito)
	○ Unique Probe based on user's defined organism 瀏覽...
Sequence (s) Paste or File upload*	>ABL AAGGTAGCTGATTTTGGCCTGAGCAGGTTGATGA CAGGGGACACCTACACAGCCCATGCTGGAGCCA AGTTCCCCA 瀏覽... DEMO
Probe Length	70
Probe # for each sequence	1 (maximum 3)
Job note (optional)	PTP family
E-mail*	yamatolin@gmail.com
Advanced Options	
[Salt]	salt_conc 0.58 (0~1M)
Degenerate probe allowed	○ yes ● no

Jobs Accepted by UPS



Dear Sir,

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

http://array.iis.sinica.edu.tw/ups/result.php?ID=20070827232034

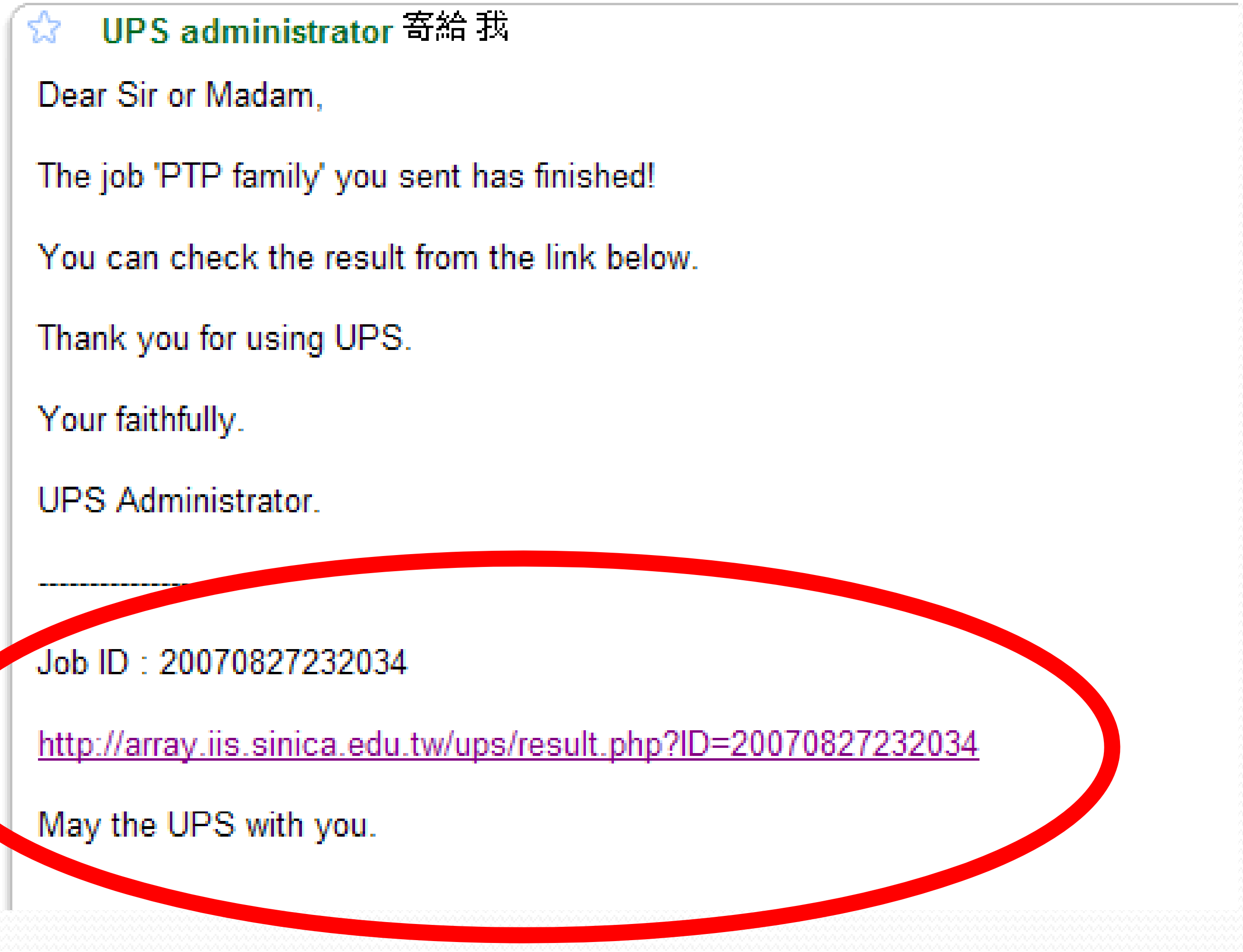
Thanks for using UPS. Any comment will be appreciated.

Your faithfully.

UPS Administrator.

Notification by Email

Message from UPS , time stamp : 2007/08/27 - 23:21:22



UPS administrator 寄給 我

Dear Sir or Madam,

The job 'PTP family' you sent has finished!

You can check the result from the link below.

Thank you for using UPS.

Your faithfully.

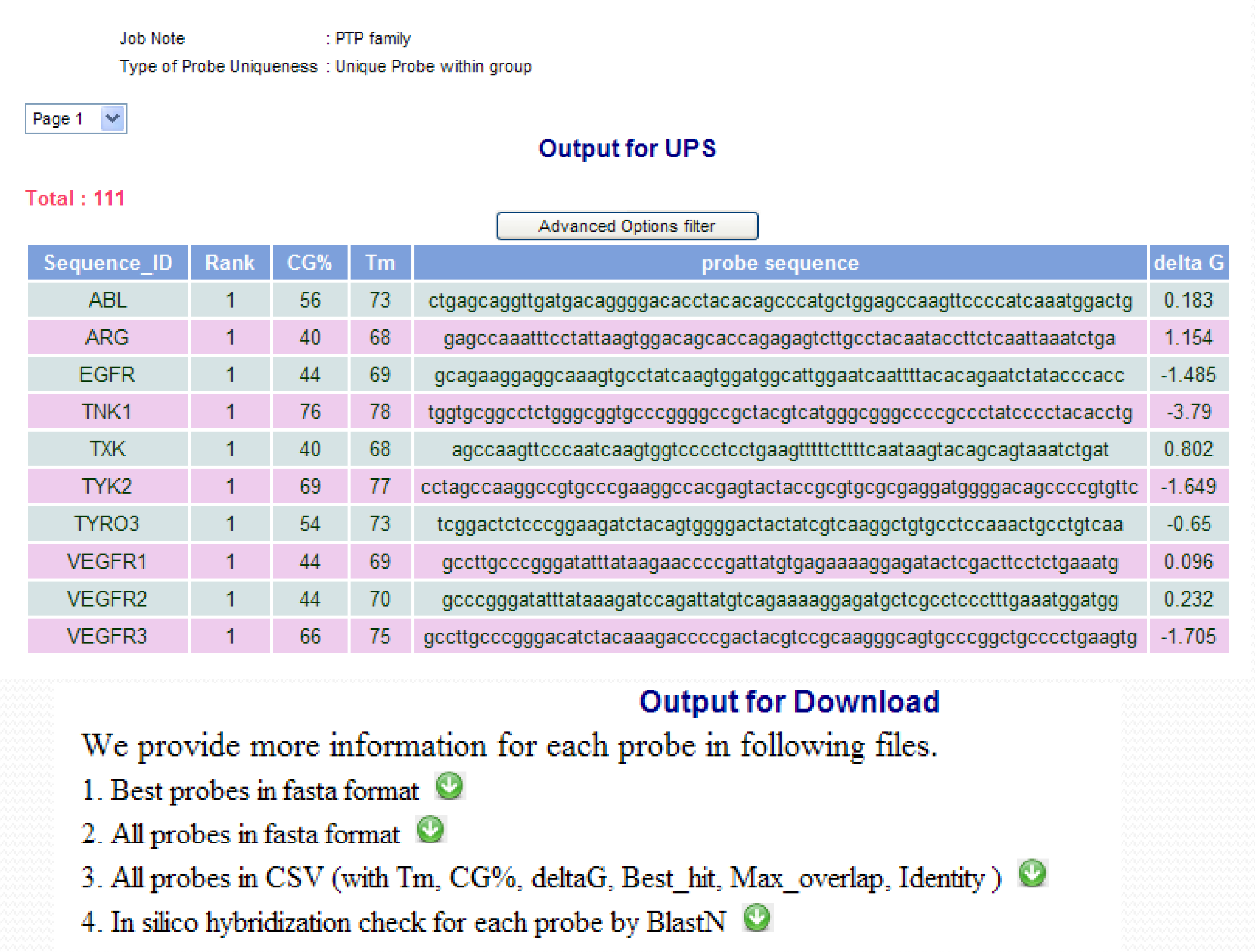
UPS Administrator.

Job ID : 20070827232034

http://array.iis.sinica.edu.tw/ups/result.php?ID=20070827232034

May the UPS with you.

Output of UPS



Job Note : PTP family

Type of Probe Uniqueness : Unique Probe within group

Output for UPS

Total : 111

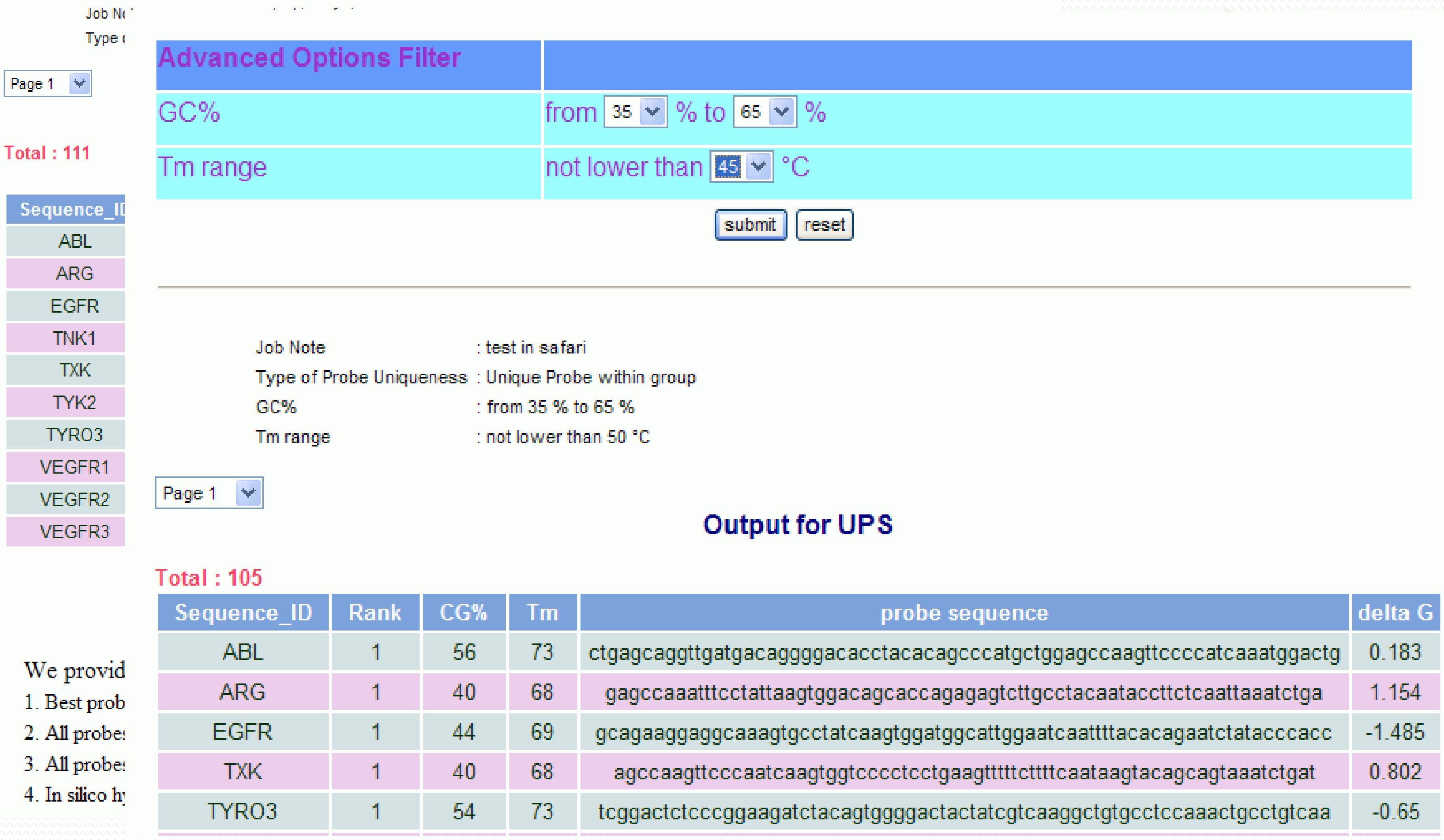
Sequence_ID	Rank	CG%	Tm	probe sequence	delta G
ABL	1	56	73	ctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatggactg	0.183
ARG	1	40	68	gagccaaatttcctattaagtggacagcaccagagagtcttgcctacaataccttctcaattaaatctga	1.154
EGFR	1	44	69	gcagaaggaggcaaagtgcctatcaagtggatggcattggaatcaattttacacagaatctatacccacc	-1.485
TNK1	1	76	78	tggtgcggcctctgggcggtgcccggggccgctacgtcatgggcgggccccgccctatcccctacacctg	-3.79
TXK	1	40	68	agccaagttcccaatcaagtggtcccctcctgaagtttttcttttcaataagtacagcagtaaatctgat	0.802
TYK2	1	69	77	cctagccaaggccgtgcccgaaggccacgagtactaccgcgtgcgcgaggatggggacagccccgtgttc	-1.649
TYRO3	1	54	73	tcggactctcccggaagatctacagtggggactactatcgtcaaggctgtgcctccaaactgcctgtcaa	-0.65
VEGFR1	1	44	69	gccttgcccgggatatttataagaaccccgattatgtgagaaaaggagatactcgacttcctctgaaatg	0.096
VEGFR2	1	44	70	gcccgggatatttataaagatccagattatgtcagaaaaggagatgctcgcctccctttgaaatggatgg	0.232
VEGFR3	1	66	75	gccttgcccgggacatctacaaagaccccgactacgtccgcaagggcagtgcccggctgcccctgaagtg	-1.705

Output for Download

We provide more information for each probe in following files.

1. Best probes in fasta format
2. All probes in fasta format
3. All probes in CSV (with Tm, CG%, deltaG, Best_hit, Max_overlap, Identity)
4. In silico hybridization check for each probe by BlastN

Advanced Options



Job N
Type

Page 1

Total : 111

Sequence_ID
ABL
ARG
EGFR
TNK1
TXK
TYK2
TYRO3
VEGFR1
VEGFR2
VEGFR3

We provid
1. Best prob
2. All probe
3. All probe
4. In silico hy

Advanced Options Filter	
GC%	from 35 % to 65 %
Tm range	not lower than 45 °C

Job Note : test in safari
Type of Probe Uniqueness : Unique Probe within group
GC% : from 35 % to 65 %
Tm range : not lower than 50 °C

Page 1

Output for UPS

Total : 105

Sequence_ID	Rank	CG%	Tm	probe sequence	delta G
ABL	1	56	73	ctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagttccccatcaaatggactg	0.183
ARG	1	40	68	gagccaaatttcctattaagtggacagcaccagagagtcttgcctacaataccttctcaattaaatctga	1.154
EGFR	1	44	69	gcagaaggaggcaaagtgcctatcaagtggatggcattggaatcaattttacacagaatctatacccacc	-1.485
TXK	1	40	68	agccaagttcccaatcaagtggtcccctcctgaagtttttcttttcaataagtacagcagtaaatctgat	0.802
TYRO3	1	54	73	tcggactctcccggaagatctacagtggggactactatcgtcaaggctgtgcctccaaactgcctgtcaa	-0.65

Output for Download

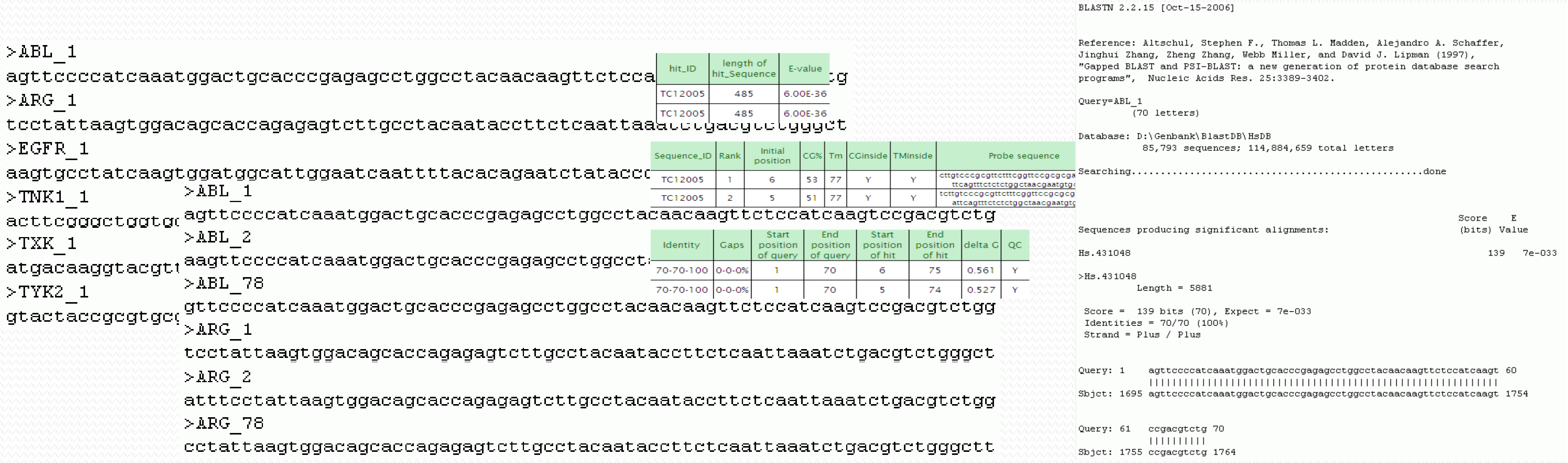
We provide more information for each probe in following files.

① Best probes in fasta format

② All probes in fasta format

③ All probes in CSV (with Tm, CG%, deltaG, Best_hit, Max_overlap, Identity)

④ In silico hybridization check for each probe by BlastN



```
>ABL_1
agttccccatcaaatggactgcacccgagagcctggcctacaacaagttctcca
>ARG_1
tcctattaagtggacagcaccagagagtcttgcctacaatacctttctcaattaa
>EGFR_1
aagtgcctatcaagtggatggcattggaatcaattttacacagaatctatacc
>TNK1_1
acttcgggctggtg
>TXK_1
atgacaaggtacgt
>TYK2_1
gtactaccgcgtgc
```

```
>ABL_1
agttccccatcaaatggactgcacccgagagcctggcctacaacaagttctccatcaagtccgacgtctg
>ABL_2
aagttccccatcaaatggactgcacccgagagcctggcct
>ABL_78
gttccccatcaaatggactgcacccgagagcctggcctacaacaagttctccatcaagtccgacgtctgg
>ARG_1
tcctattaagtggacagcaccagagagtcttgcctacaatacctttctcaattaaatctgacgtctgggct
>ARG_2
atttcctattaagtggacagcaccagagagtcttgcctacaatacctttctcaattaaatctgacgtctgg
>ARG_78
cctattaagtggacagcaccagagagtcttgcctacaatacctttctcaattaaatctgacgtctgggctt
```

hit_ID	length of hit_Sequence	E-value
TC12005	485	6.00E-36
TC12005	485	6.00E-36

Sequence_ID	Rank	Initial position	CG%	Tm	CGinside	TMinside	Probe sequence
TC12005	1	6	53	77	Y	Y	cttgtcccgcgttctttcggttccgcgcga ttcagtttctctctggctaacgaatgtg
TC12005	2	5	51	77	Y	Y	tcttgtcccgcgttctttcggttccgcgcg attcagtttctctctggctaacgaatgtg

Identity	Gaps	Start position of query	End position of query	Start position of hit	End position of hit	delta G	QC
70-70-100	0-0-0%	1	70	6	75	0.561	Y
70-70-100	0-0-0%	1	70	5	74	0.527	Y

```
BLASTN 2.2.15 [Oct-15-2006]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs",  Nucleic Acids Res. 25:3389-3402.

Query=ABL_1
         (70 letters)

Database: D:\Genbank\BlastDB\HsDB
           85,793 sequences; 114,884,659 total letters

Searching..................................................done



                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

Hs.431048                                                           139   7e-033

>Hs.431048
          Length = 5881

 Score =  139 bits (70), Expect = 7e-033
 Identities = 70/70 (100%)
 Strand = Plus / Plus


Query: 1    agttccccatcaaatggactgcacccgagagcctggcctacaacaagttctccatcaagt 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1695 agttccccatcaaatggactgcacccgagagcctggcctacaacaagttctccatcaagt 1754


Query: 61   ccgacgtctg 70
            ||||||||||
Sbjct: 1755 ccgacgtctg 1764
```

upQPCR :: PDA+ UPS

- For specific sequence wanted to identify by Q-PCR, the following steps can be used to get the primer pairs and probe
  1. Submit Sequence to PDA for best primer set with specific region (or select by PDA)
  2. Submit the amplicon to UPS and choose the organism you used to get the best probe for Q-PCR

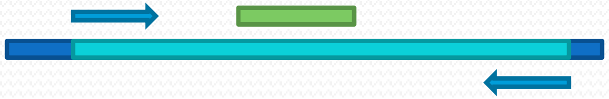
PDA
Primer Design Assistant

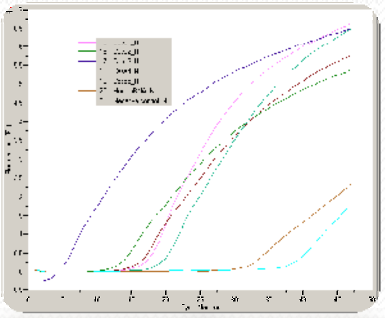
+

UPS
Unique Probe Selector

Home Demo Help Contact

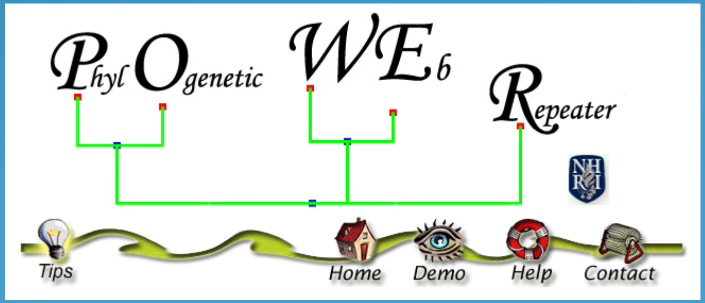
→





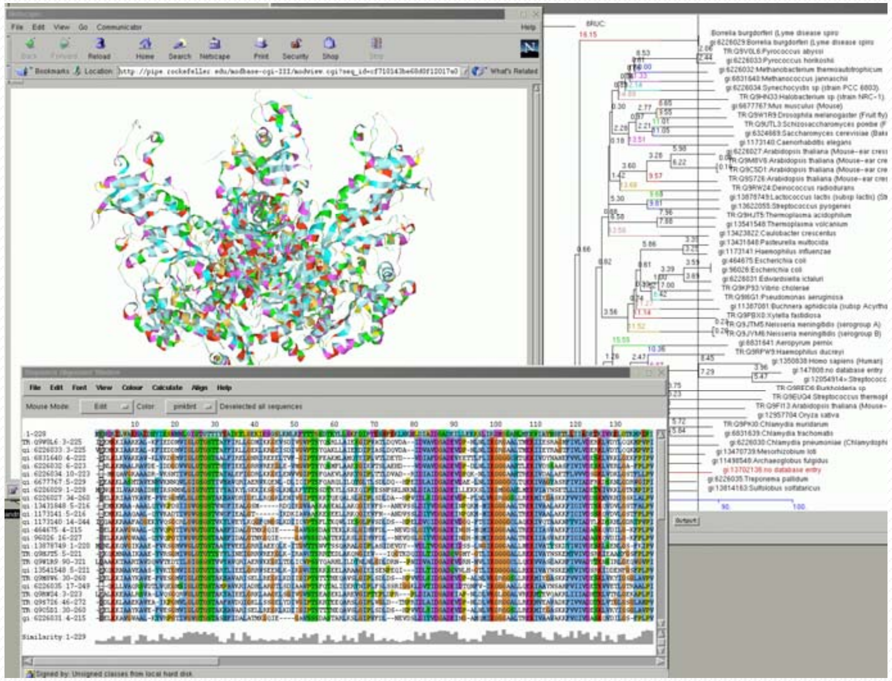
Phylogenetic Analysis

- ✓ Phylogenetic Web Repeater (POWER)
- ✓ Phylogenetic reconstruction by Automatic Likelihood Model selector (PALM)



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

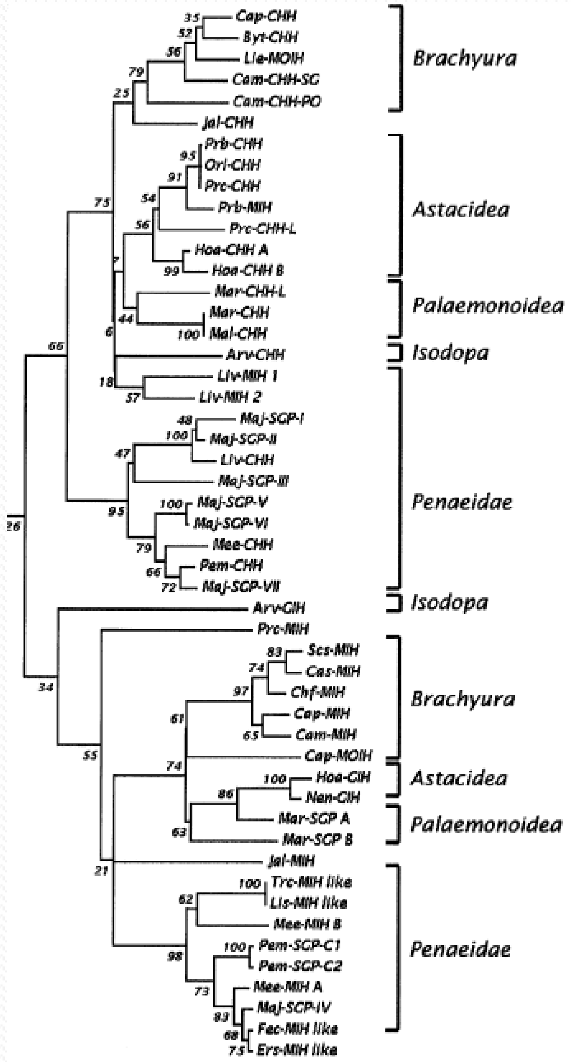
Coding Characters and Defining Homology

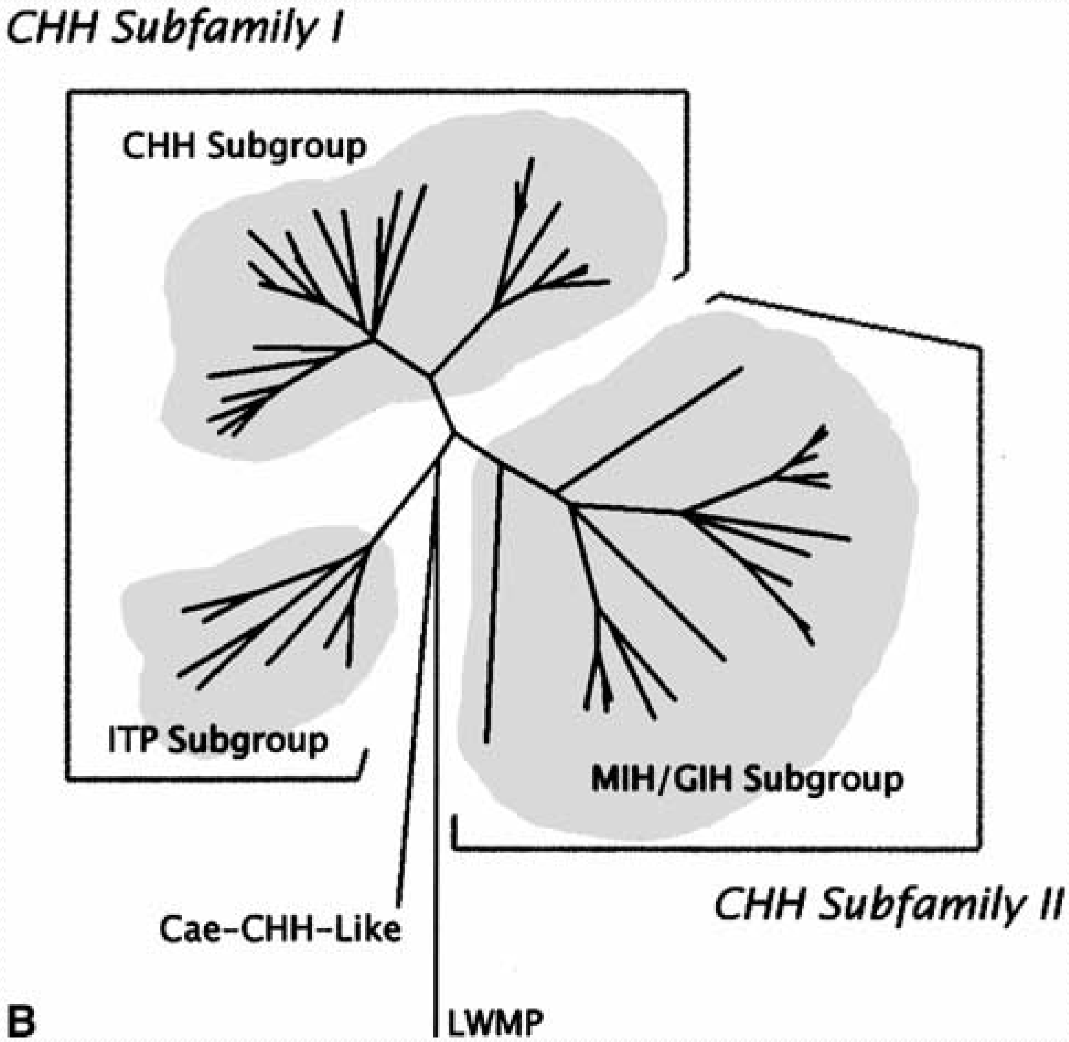


Classical phylogenetic analysis by Morphology

Molecular phylogenetic analysis By Bio-Molecules

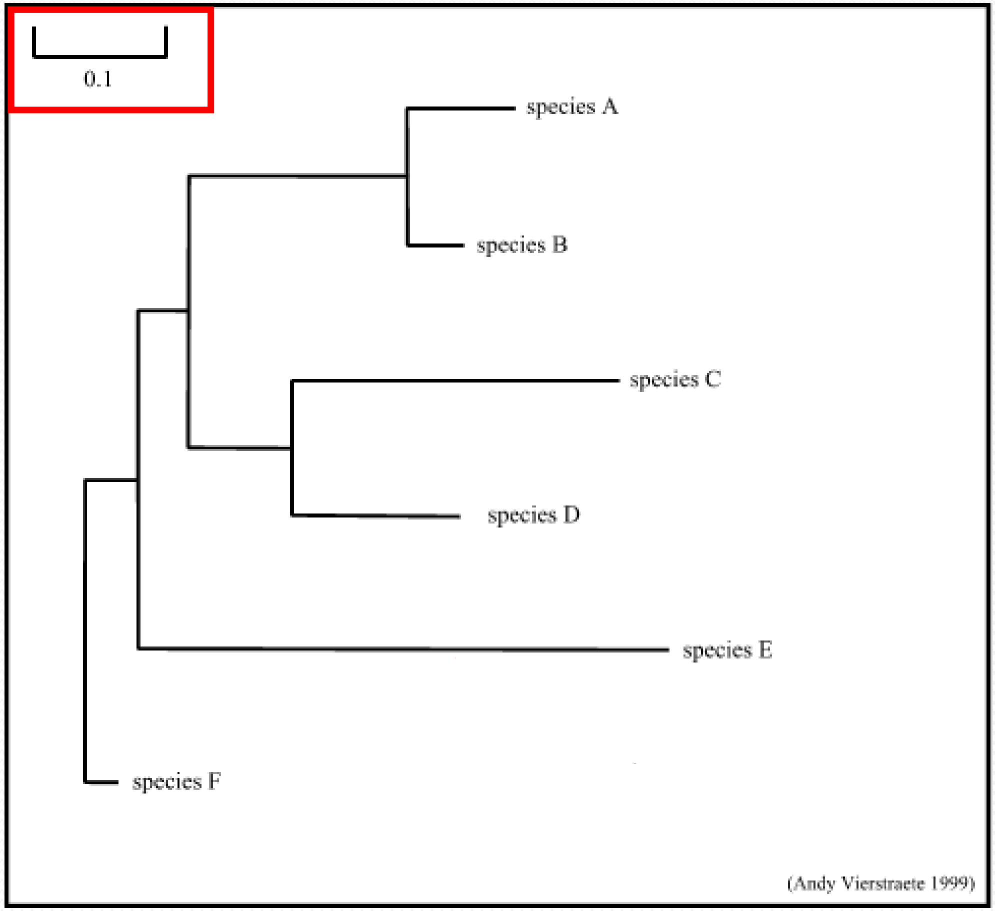
An Example of Phylogenetic Tree





Phylogenetic Tree

- The tree is composed of nodes connected by branches.

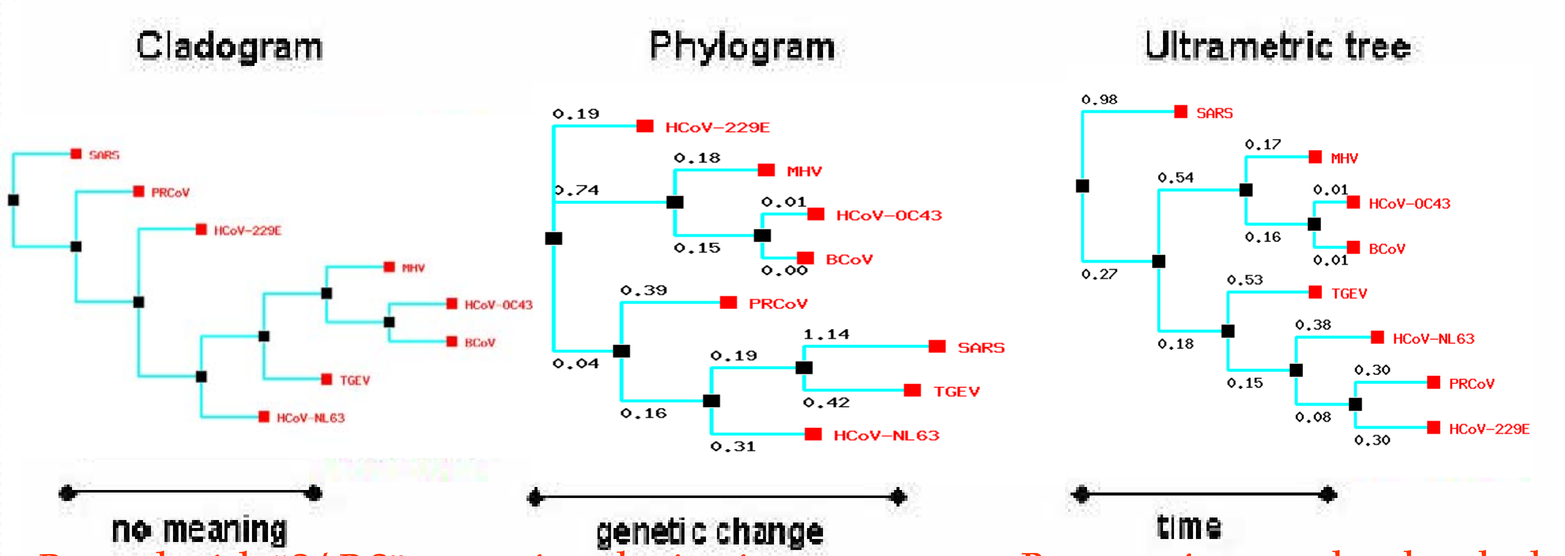


(Andy Vierstraete 1999)

- **node :** a node represents a taxonomic unit.
  - Internal nodes
  - External nodes
- **branch (edge):** defines the relationship between the taxa.
- **branch length :** often represents the number of changes that have occurred in that branch.
- **root :** is the common ancestor of all taxa.
- **distance scale :** scale which represents the number of differences between sequences (e.g. 0.1 means 10 % differences between two sequences)

Types of Phylogenetic Tree

- Branch: define relationship between nodes
  - Branch length: longer branch length, more sequence changes



Rooted with “SARS”

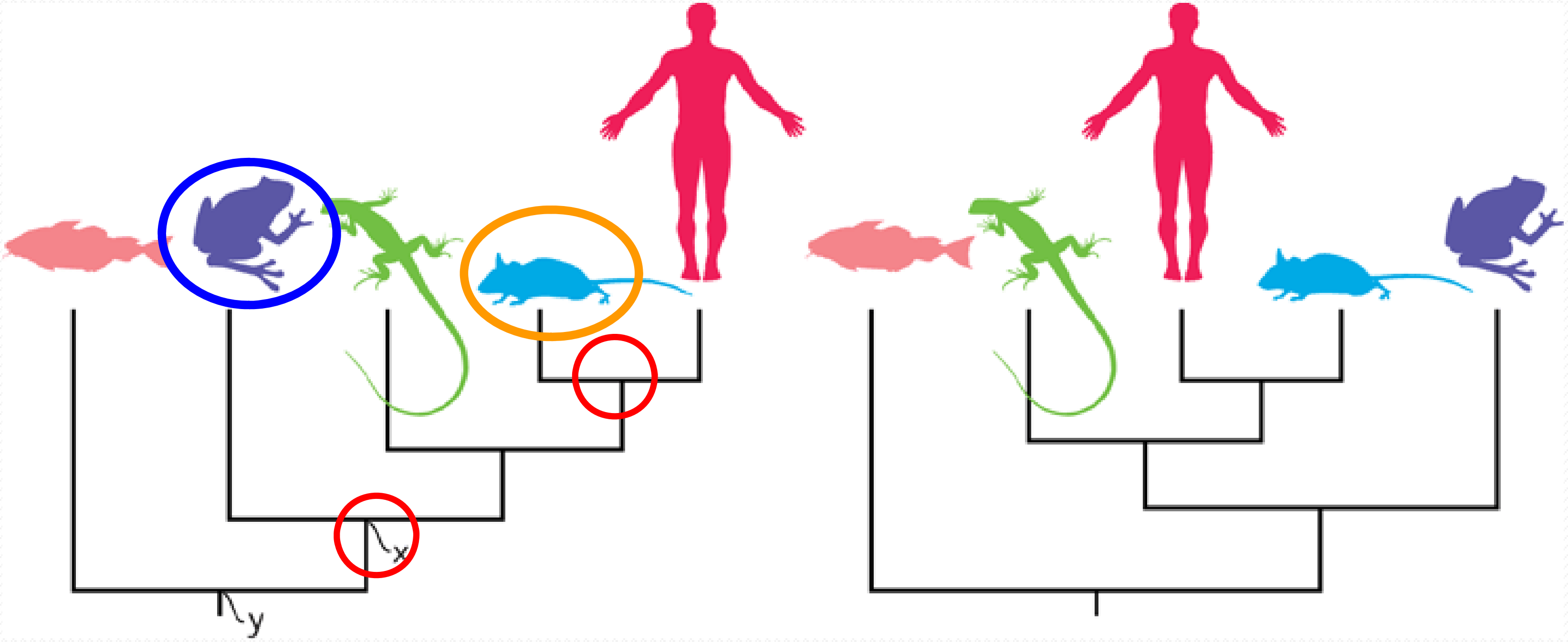
in substitutions per nucleotide

By assuming a molecular clock

Ex. Pasimony **Ex. Neighbor-join, ML** **Ex. UPGMA**

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.



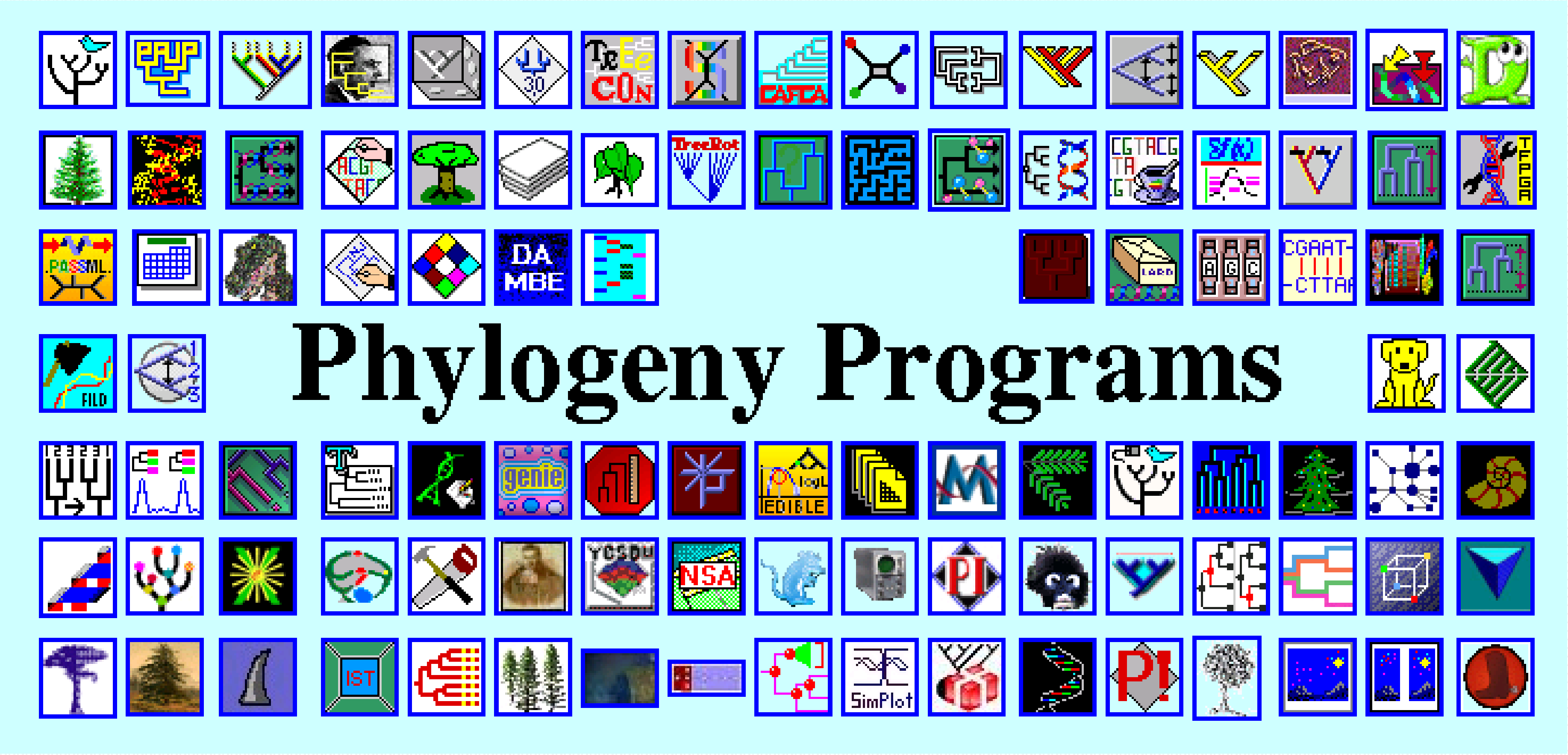
(David A. Baum et al., Science 11 November 2005:Vol. 310. no. 5750, pp. 979 – 980)

The Ways to Construct the tree

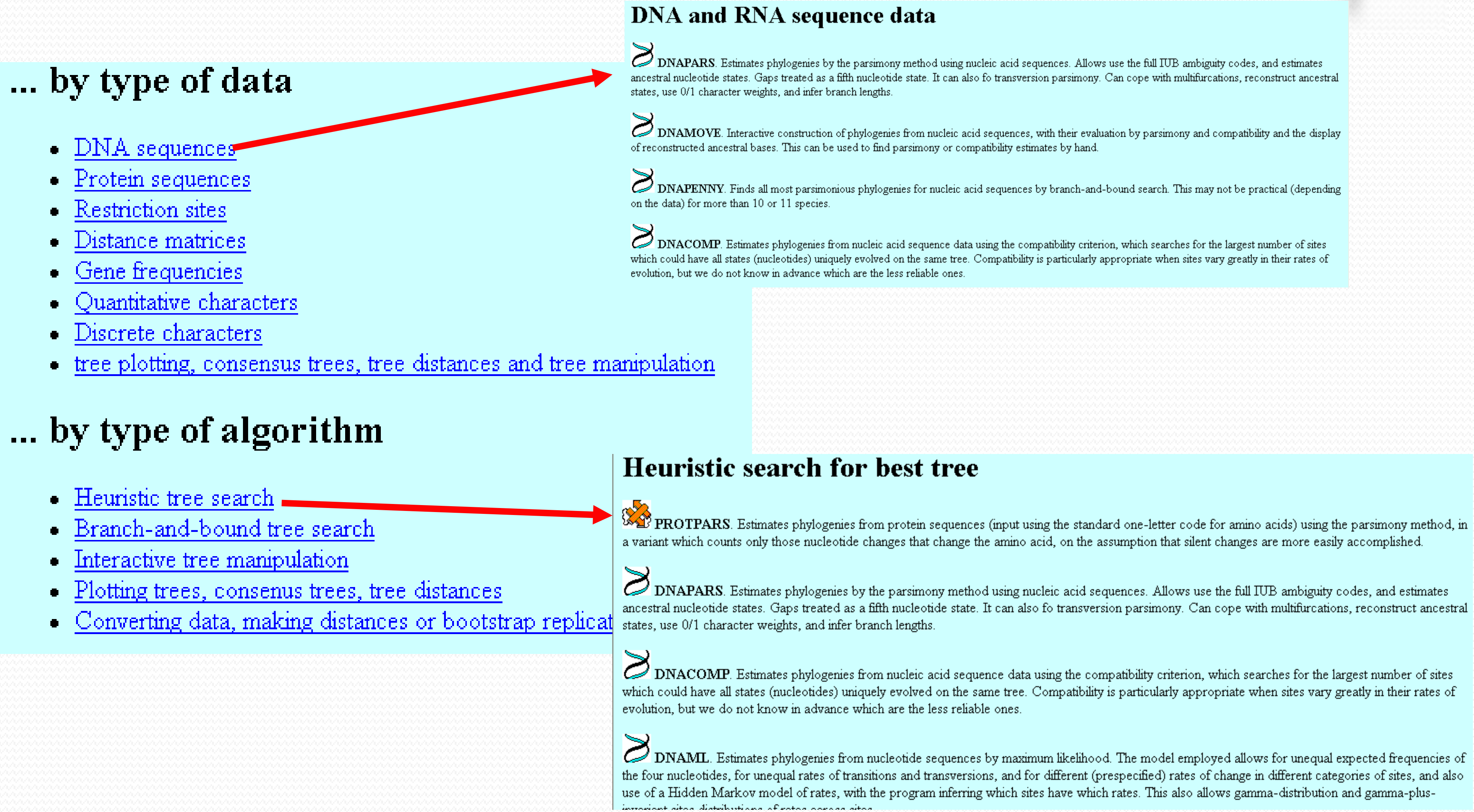
- Distance-matrix methods (Dis)
  - Neighbor-joining
  - Fitch-Margoliash method
  - Using outgroups
- Maximum parsimony (MP)
  - Branch and bound
  - Sankoff-Morel-Cedergren algorithm
  - MALIGN and POY
- Maximum likelihood (ML)
- Bayesian inference (BI)

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

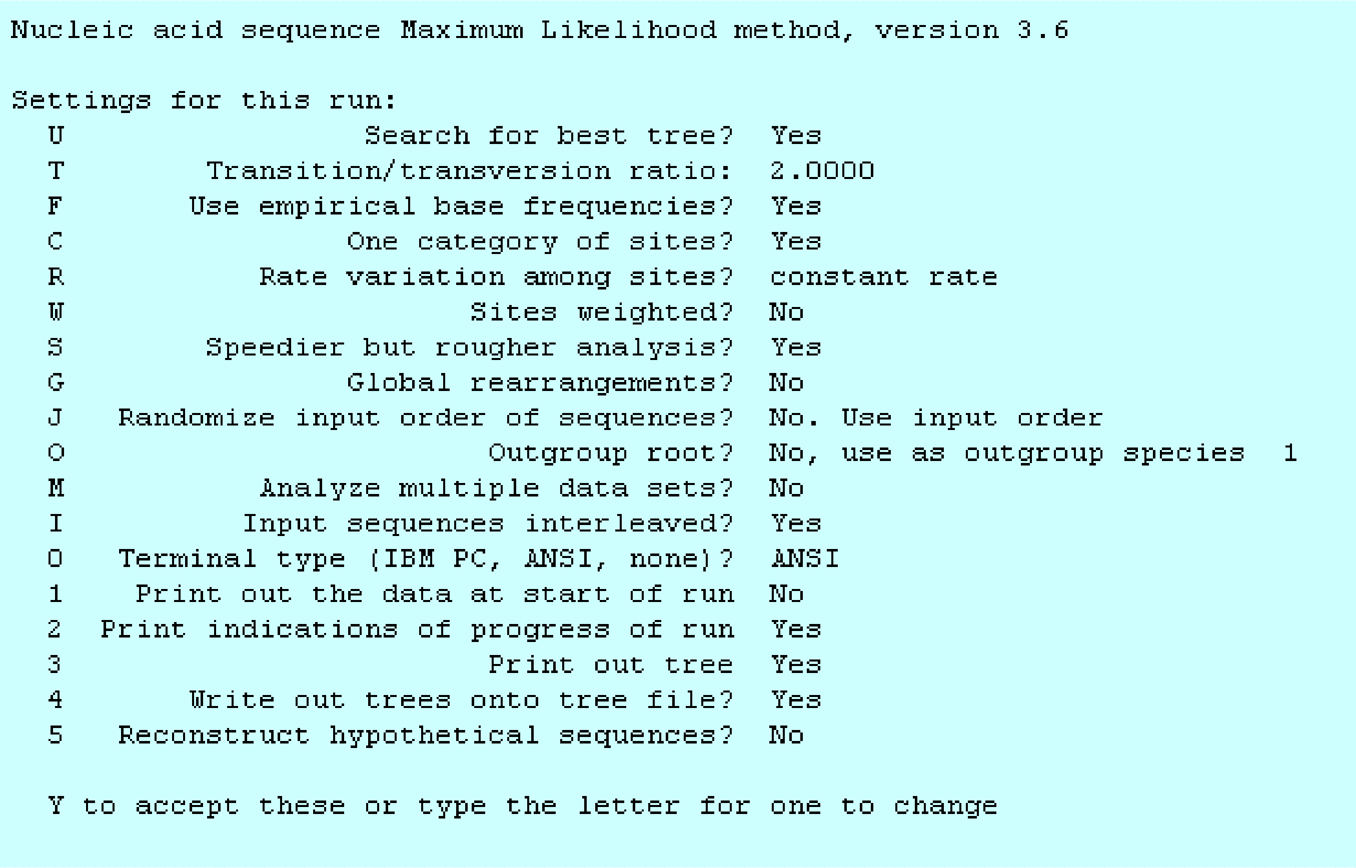
PROTPARS. Estimates phylogenies from protein sequences (input using the standard one-letter code for amino acids) using the parsimony method, in a variant which counts only those nucleotide changes that change the amino acid, on the assumption that silent changes are more easily accomplished.

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



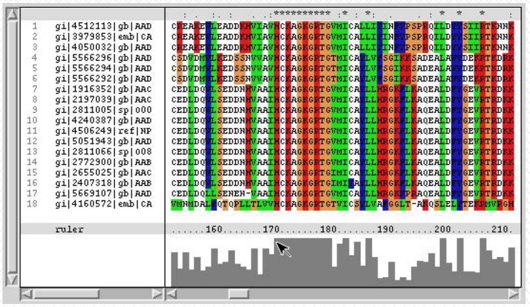
```
Nucleic acid sequence Maximum Likelihood method, version 3.6

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 Analysis

- Character state method
  - Maximum parsimony
- Distance method
  - Neighbor-joining and UPGMA method
  - Fitch-Margoliash method
- Maximum likelihood methods
  - determinate evolution model first, then construct system trees

General Pipeline for Phylogenetic Analysis

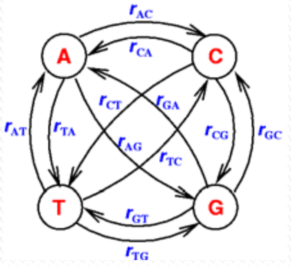


```
                              :   .  : :  :   .  . :*********** :* : *:  .   . : *  *   *  :
 1   gi|4512113|gb|AAD    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 2   gi|3979853|emb|CA    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 3   gi|4050032|gb|AAD    CREAKEWLEADDKHVIAVHCKAGKGRTGVMICALLIYINFYPSPRQILDYYSIIRTKNNK
 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    CEDLDQWLSEDDNHVAAIHCKAGKGRTGIMICAYLLHRGKFLKAQEALDFYGEVRTRDKK
17   gi|5669107|gb|AAD    CEDLDQLLSENEN-VAAIHCKAGKGRTGVMICAYLLHRGKFPRAQEALDFYGEVRTRDKK
18   gi|4160572|emb|CA    VMNMDALFQTQPLLTLVVHCKAGKGRTGTVICSYLVAFGGLT-AKQSLELLTEKRMVRGH

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

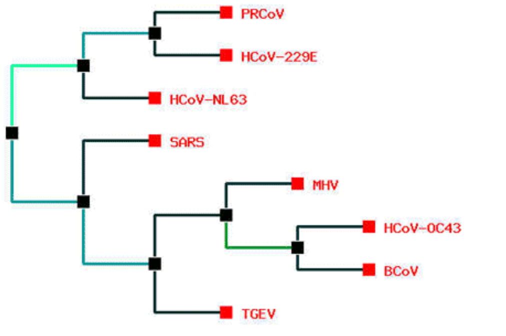
Multiple Sequence Alignment

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 mothodes	• Maximum likelihood method • Maximum likelihood method with molecular clock	

Selection of inference Methods

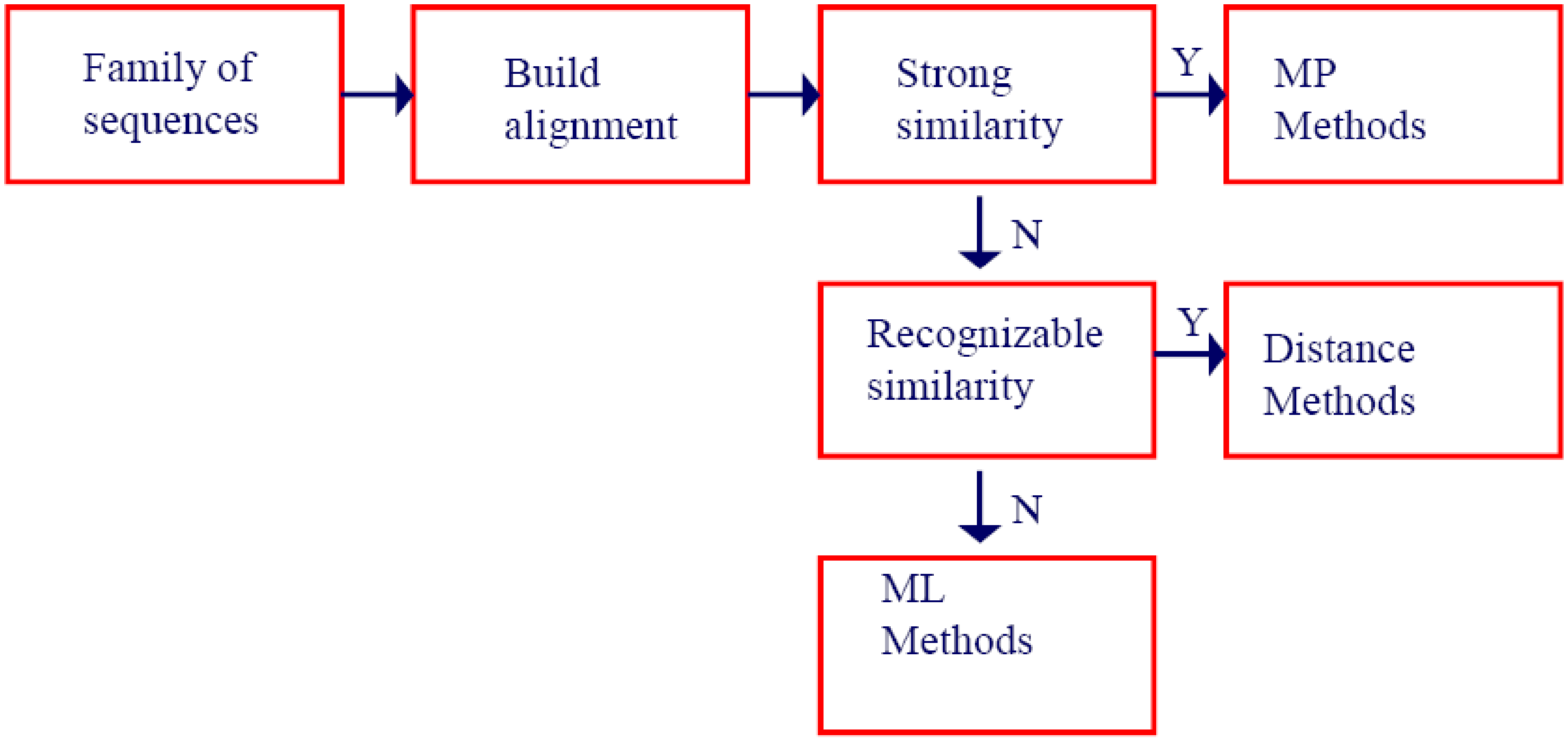


Bootstrap
Substitution Model
Tree Construction



Evaluate phylogenetic tree

General Rule for Method Selection



Family of sequences

Build alignment

Strong similarity

Y

MP Methods

N

Recognizable similarity

Y

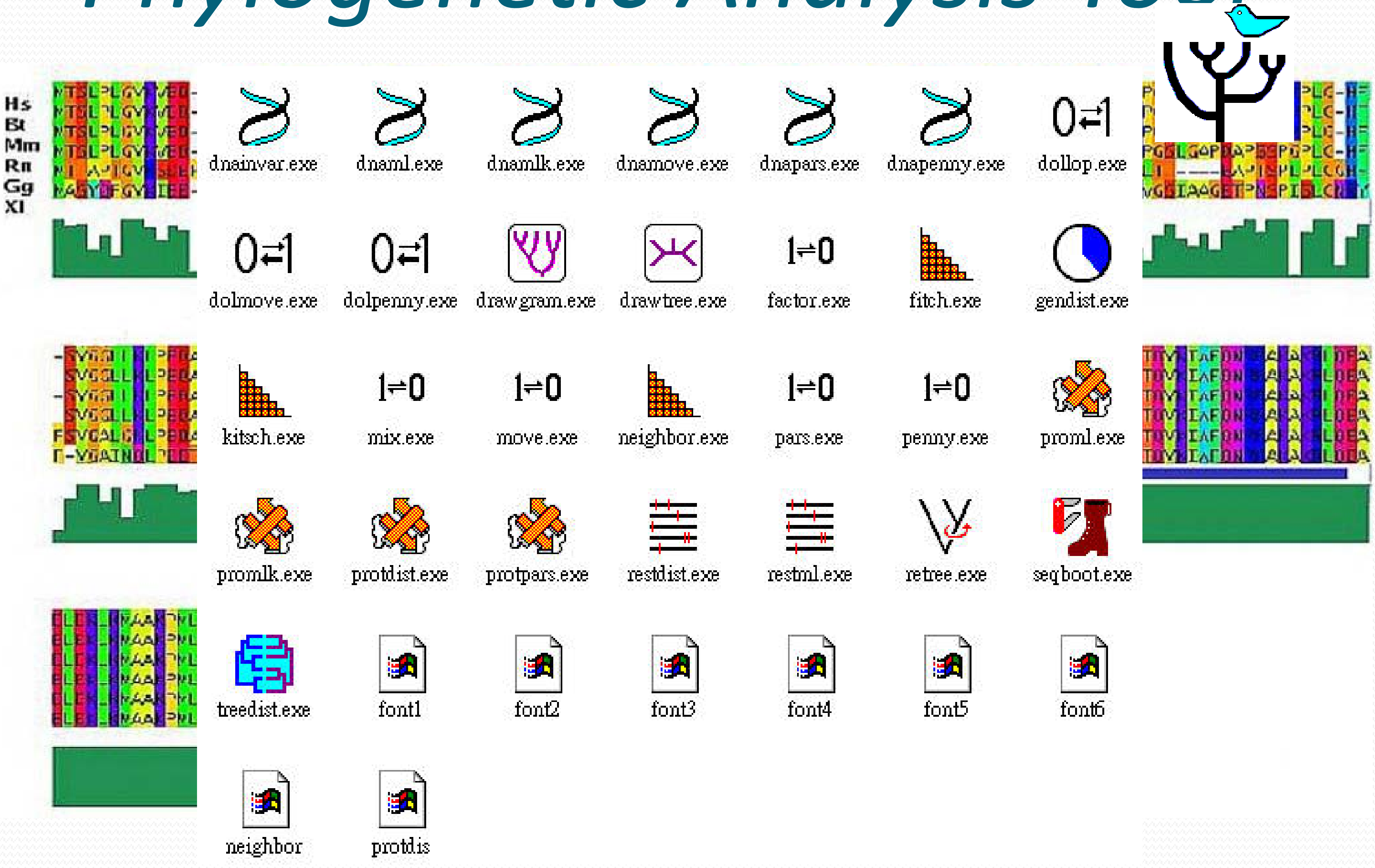
Distance Methods

N

ML Methods

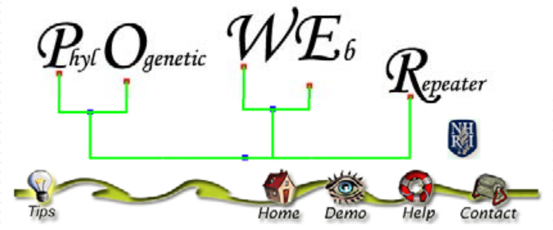
(Mount, *Bioinformatics*)

Phylogenetic Analysis Tool

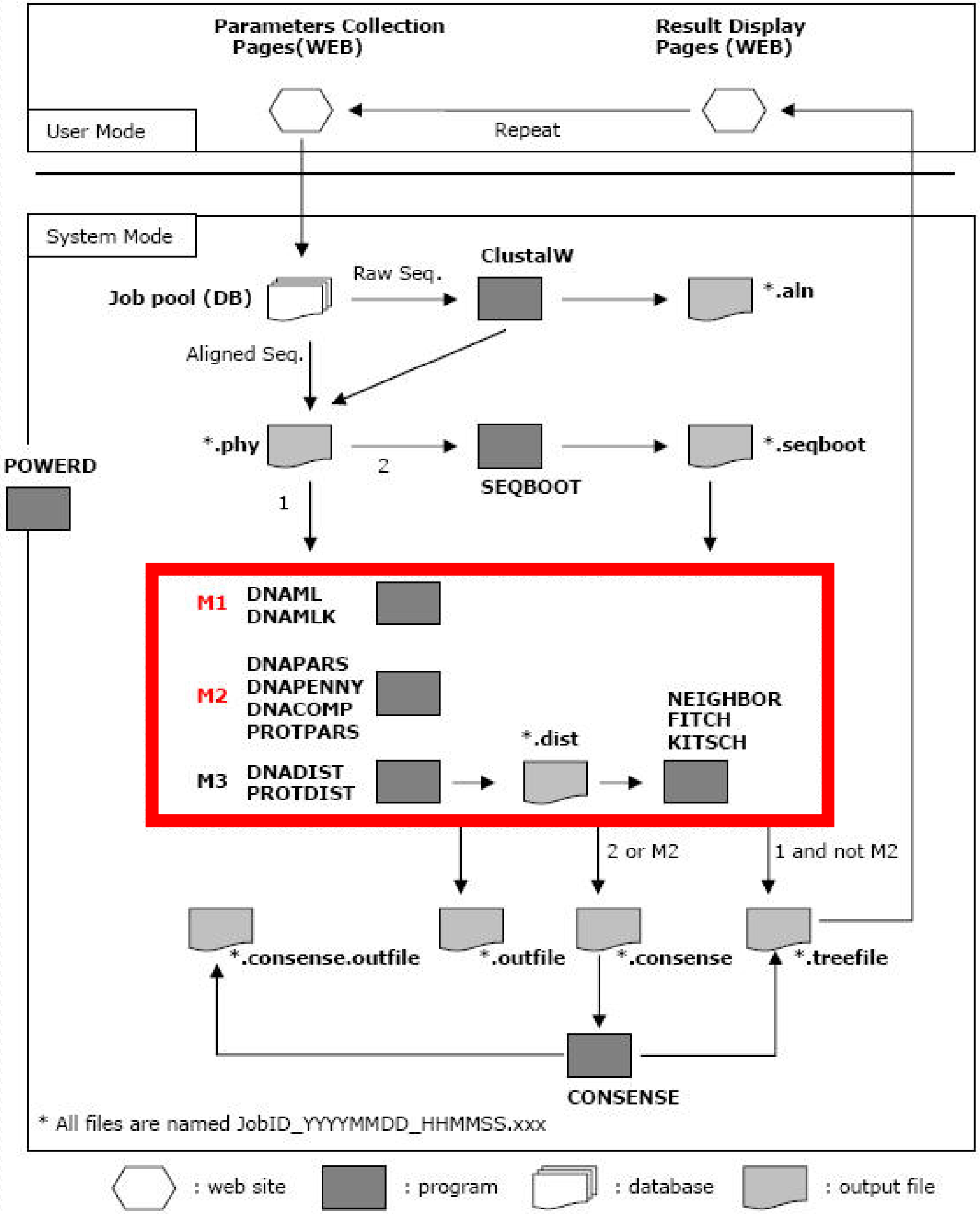


POWER: PhylOgenetic WEb Repeater

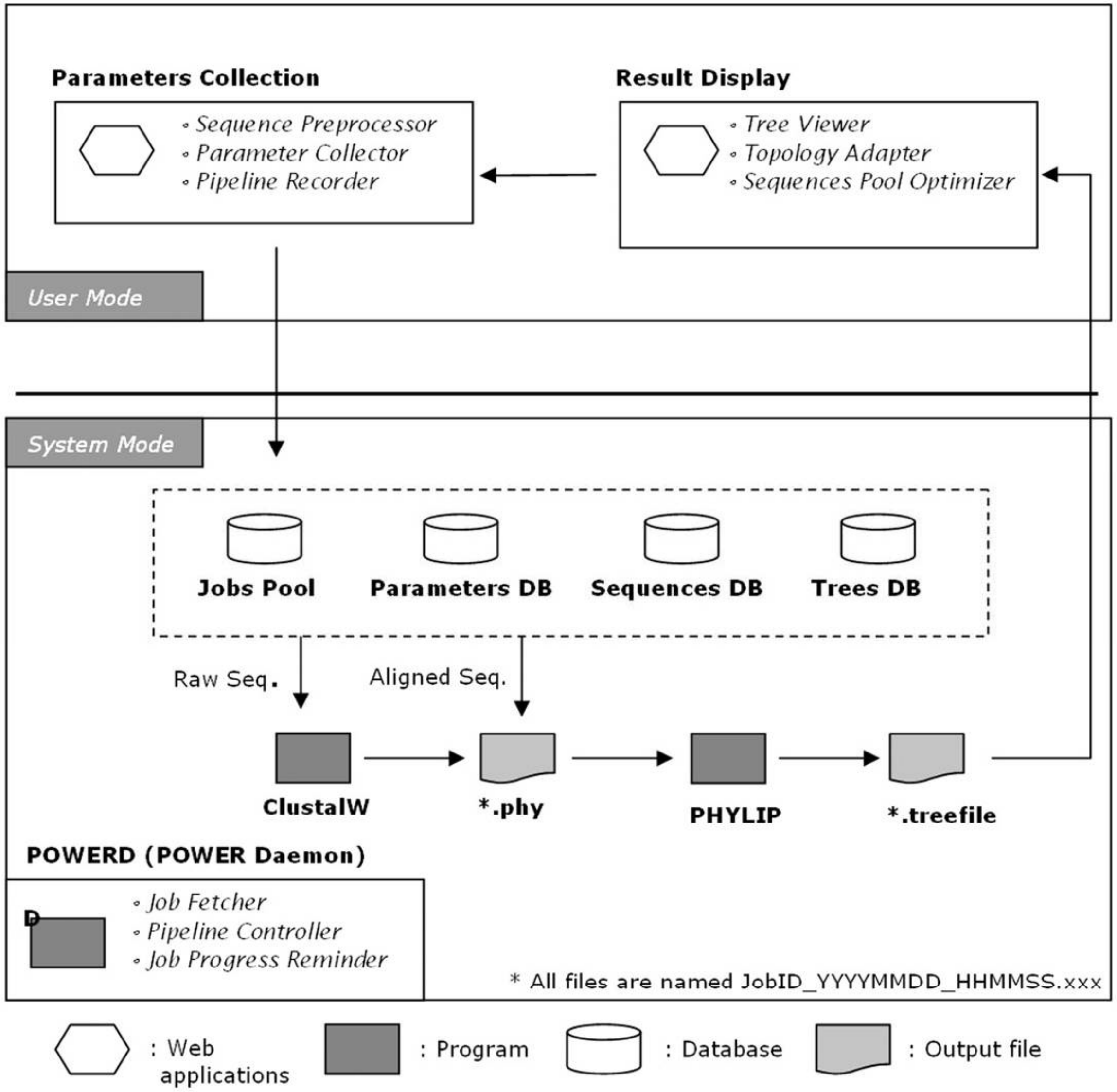
- Provide a seamless way to conduct the complex phylogenetic analysis for Biologists
- An integrated and user-optimized framework for biomolecular phylogenetic analysis
- POWER uses an open-source LAMP (Linux, Apache, MySQL, PHP) structure and infers genetic distances and phylogenetic relationships using 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



Make Phylip Packages into Automatic Flow

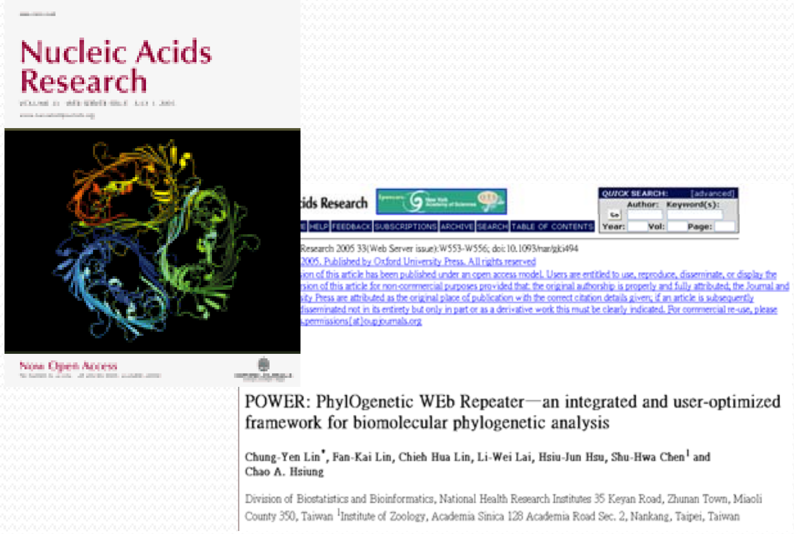


Inside of POWER



POWER: PhylOgenetic WEb Repeater

http://power.nhri.org.tw



Nucleic Acids Research

ids Research

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

2005. Published by Oxford University Press. All rights reserved

ion 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
ity 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

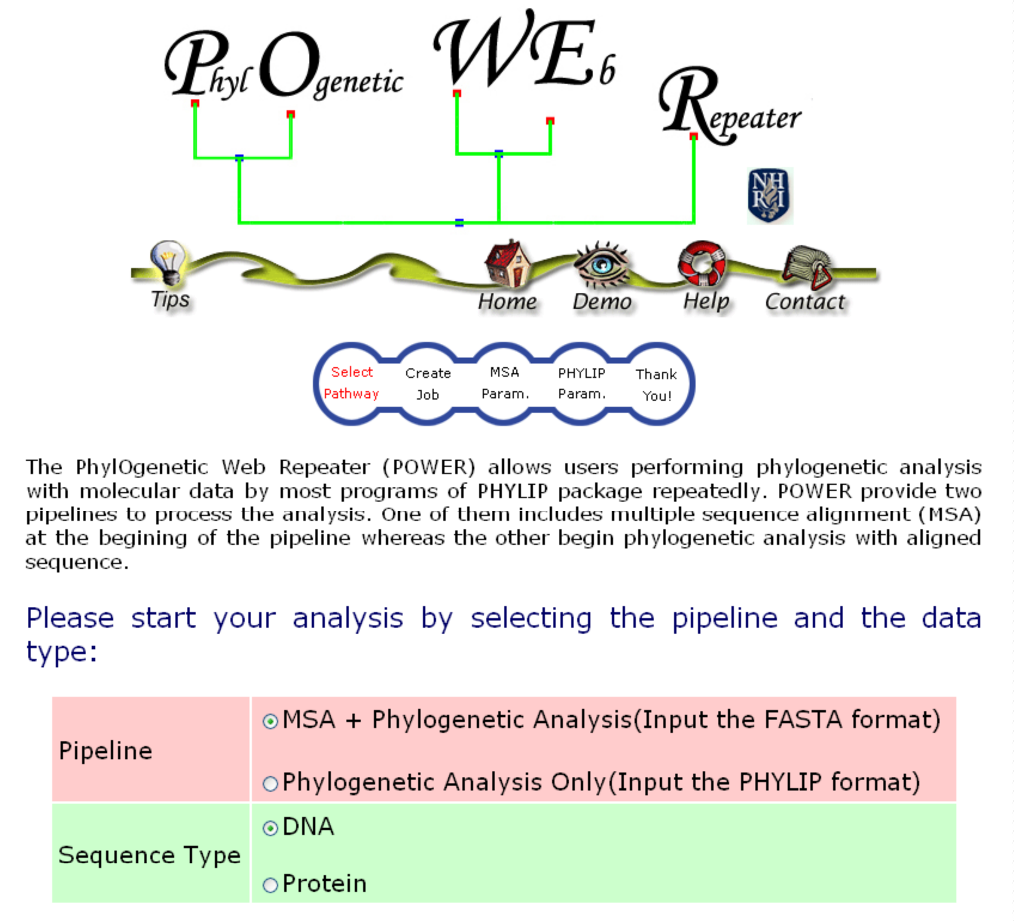
Now Open Access

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



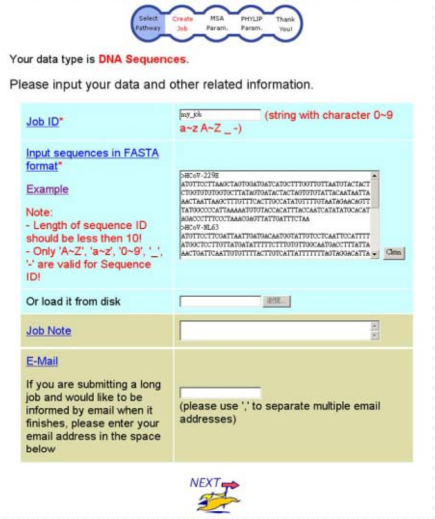
The PhylOgenetic Web Repeater (POWER) allows users performing phylogenetic analysis with molecular data by most programs of PHYLIP package repeatedly. POWER provide two pipelines to process the analysis. One of them includes multiple sequence alignment (MSA) at the begining of the pipeline whereas the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

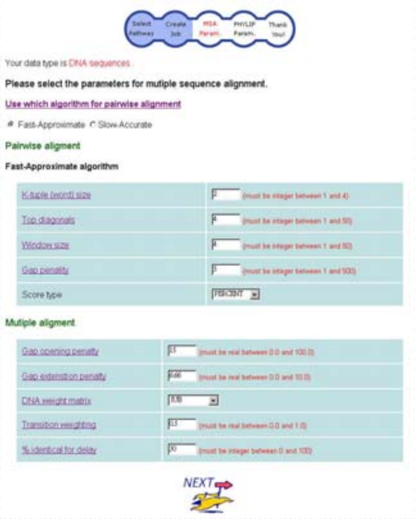


Your data type is DNA Sequences.

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

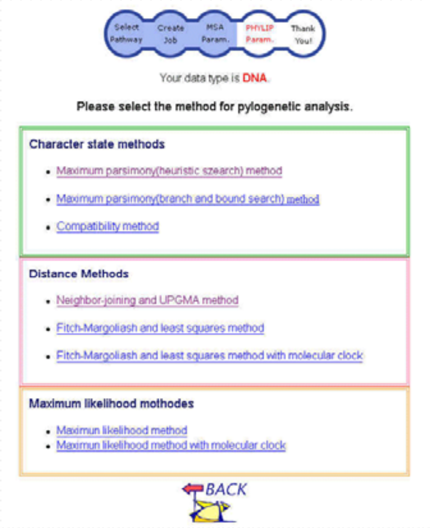
Fast-Approximate algorithm

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 penalty	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 extension 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

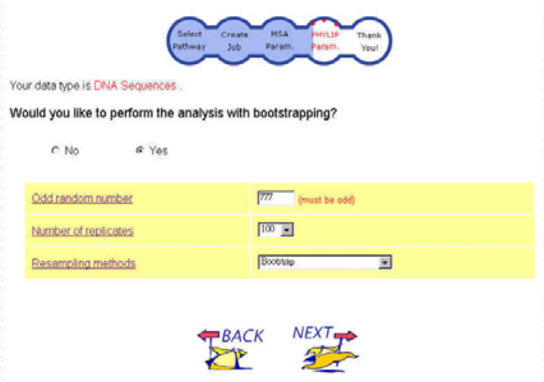
- Neighbor-joining and UPGMA method
- Fitch-Margoliash and least squares method
- Fitch-Margoliash and least squares method with molecular clock

Maximum likelihood mothodes

- Maximun likelihood method
- Maximun likelihood method with molecular clock

PhylOgenetic Web Repeater (POWER)

Options of bootstrapping



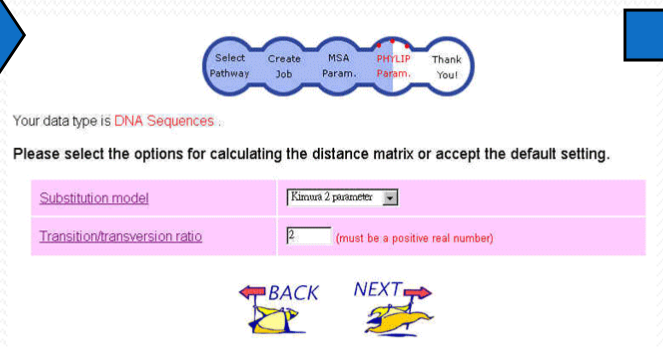
Your data type is DNA Sequences .

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

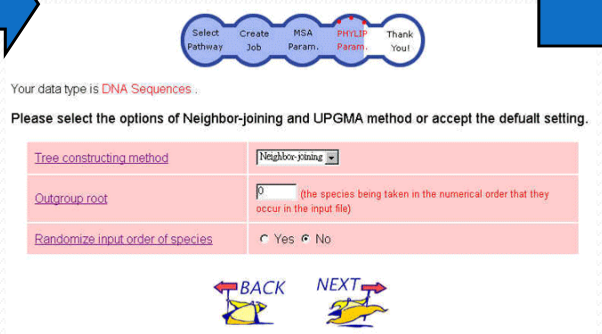


Your data type is DNA Sequences .

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 .

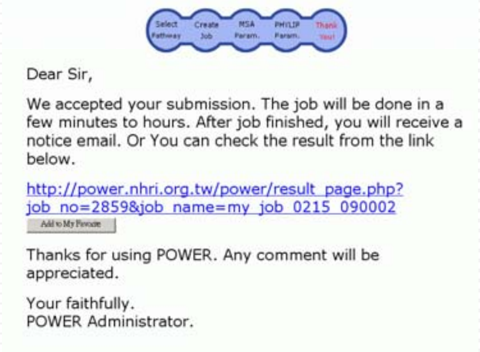
Please select the options of Neighbor-joining and UPGMA method or accept the defualt setting.

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,

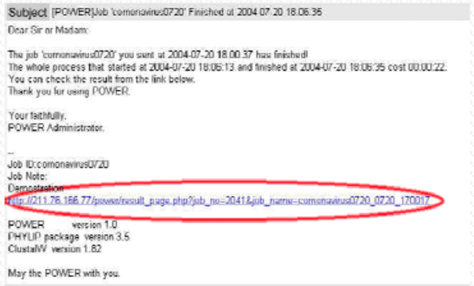
We accepted your submission. The job will be done in a few minutes to hours. After job finished, you will receive a notice email. Or You can check the result from the link below.

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 'comonavirus0720' Finished at 2004.07.20 18:06:35

Dear Sir or Madam:

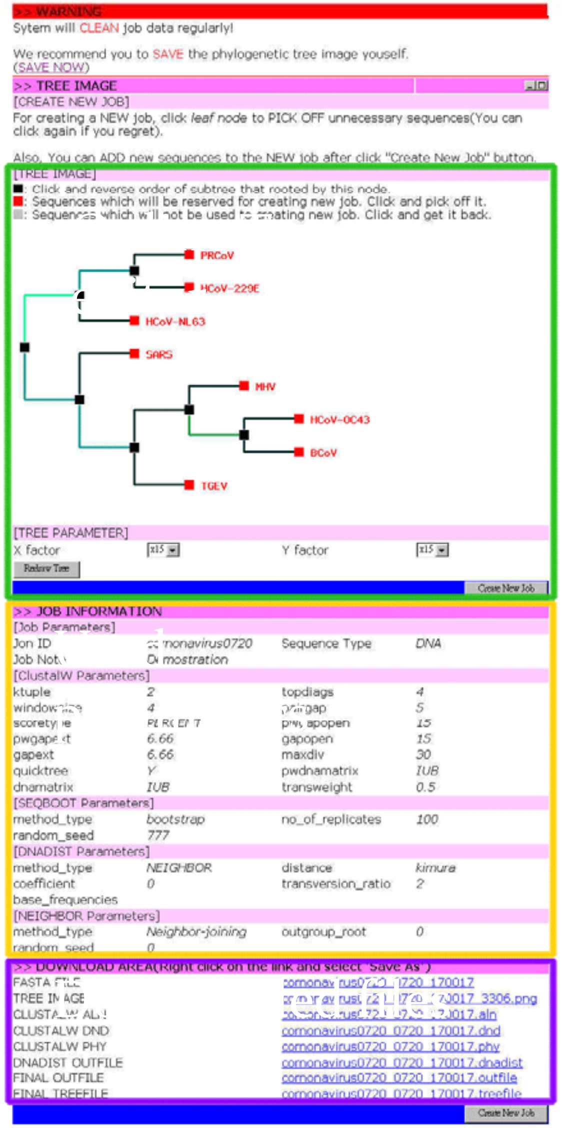
The job 'comonavirus0720' you sent at 2004-07-20 18:00:37 has finished!
The whole process that started at 2004-07-20 18:06:13 and finished at 2004-07-20 18:06:35 cost 00:00:22.
You can check the result from the link below.
Thank you for using POWER.

Your faithfully.
POWER Administrator.

--
Job ID:comonavirus0720
Job Note:
Demostration
http://211.76.166.77/power/result_page.php?job_no=2041&job_name=comonavirus0720_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]

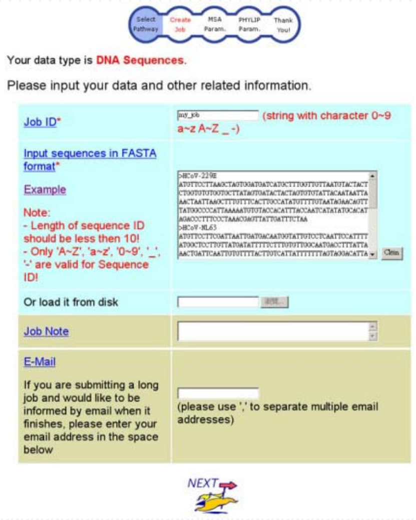
X factor x15 ▾ Y factor x15 ▾

>> JOB INFORMATION

[Job Parameters]			
Jon ID	comonavirus0720	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	comonavirus0720_0720_170017
TREE IMAGE	comonavirus0720_0720_170017_3306.png
CLUSTALW ALN	comonavirus0720_0720_170017.aln
CLUSTALW DND	comonavirus0720_0720_170017.dnd
CLUSTALW PHY	comonavirus0720_0720_170017.phy
DNADIST OUTFILE	comonavirus0720_0720_170017.dnadist
FINAL OUTFILE	comonavirus0720_0720_170017.outfile
FINAL TREEFILE	comonavirus0720_0720_170017.treefile



Your data type is **DNA Sequences**.

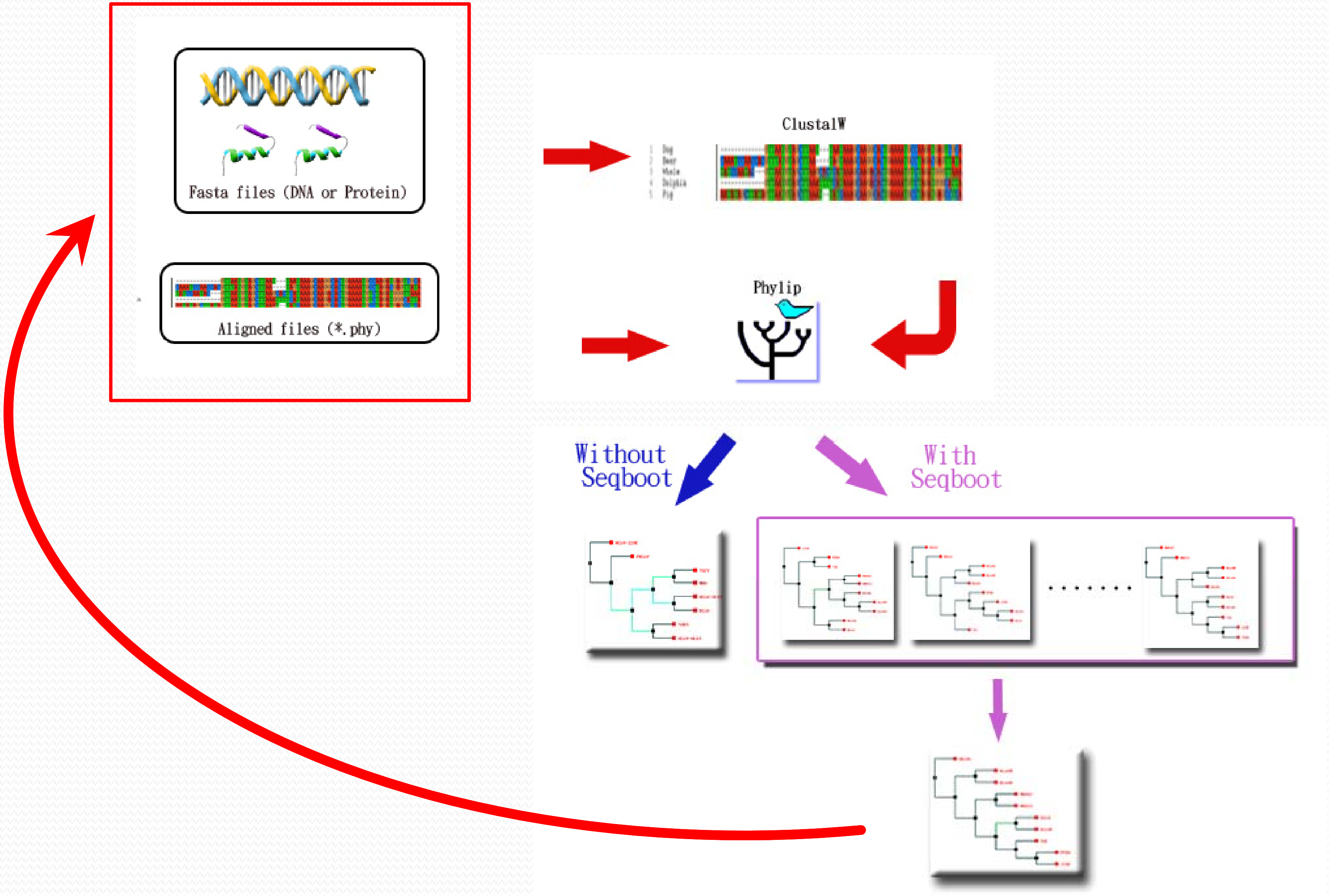
Please input your data and other related information.

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!	(see sequences below) 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)

```
>HCoV-229E
ATGTTCCTTAAGCTAGTGGATGATCATGCTTTGGTTGTTAATGTACTACT
CTGGTGTGTGGTGCTTATAGTGATACTACTAGTGTGTATTACAATAATTA
AACTAATTAAGCTTTGTTTCACTTGCCATATGTTTTGTAATAGAACAGTT
TATGGCCCCATTAAAAATGTGTACCACATTTACCAATCATATATGCACAT
AGACCCTTTCCCTAAACGAGTTATTGATTTCTAA
>HCoV-NL63
ATGTTCCTTCGATTAATTGATGACAATGGTATTGTCCTCAATTCCATTTT
ATGGCTCCTTGTTATGATATTTTTCTTTGTGTTGGCAATGACCTTTATTA
AACTGATTCAATTGTGTTTTACTTGTCATTATTTTTTTAGTAGGACATTA
```

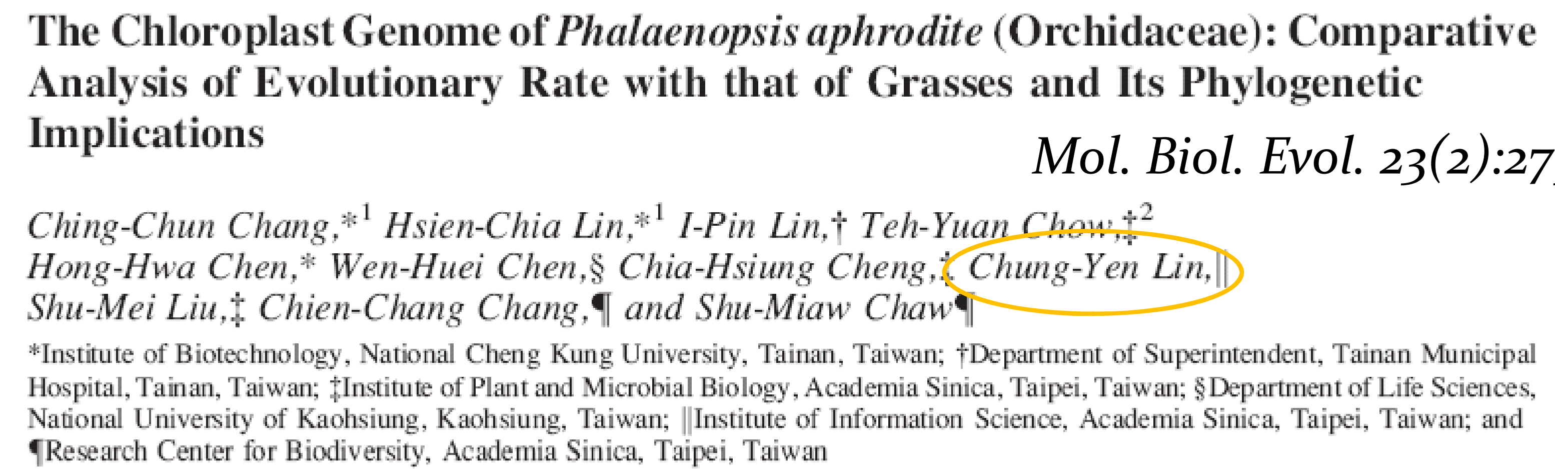
Re-perform the process by items added or deleted

PhylOgenetic Web Repeater (POWER)



Add/ delete sequences to invoke new job

Publication in POWER

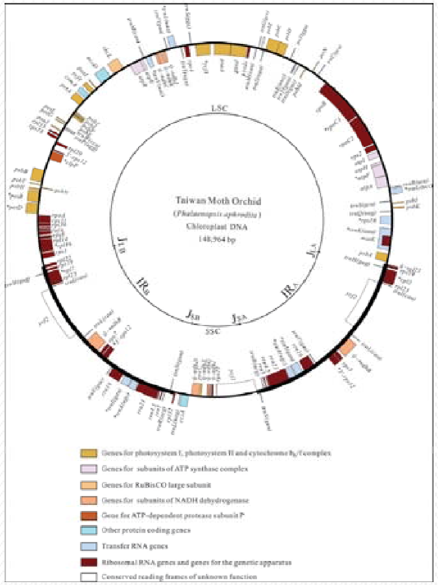


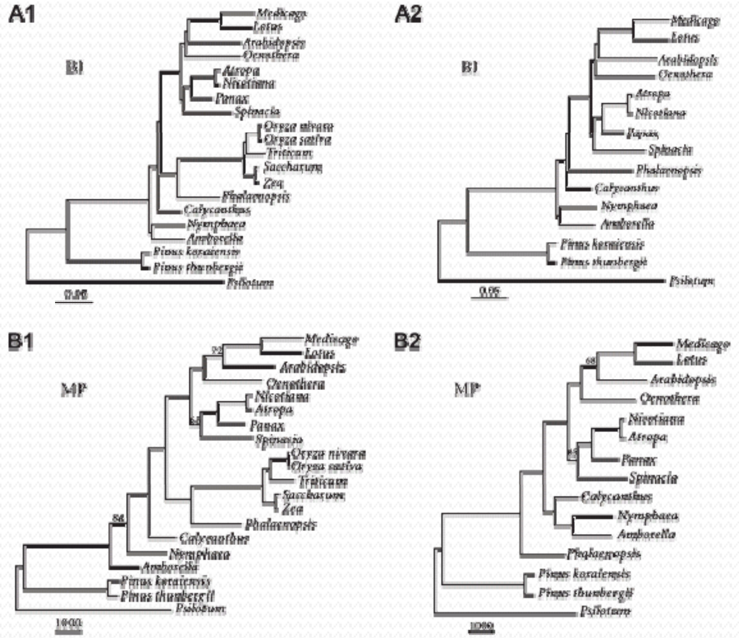
The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with that of Grasses and Its Phylogenetic Implications

Mol. Biol. Evol. 23(2):279–291. 2006

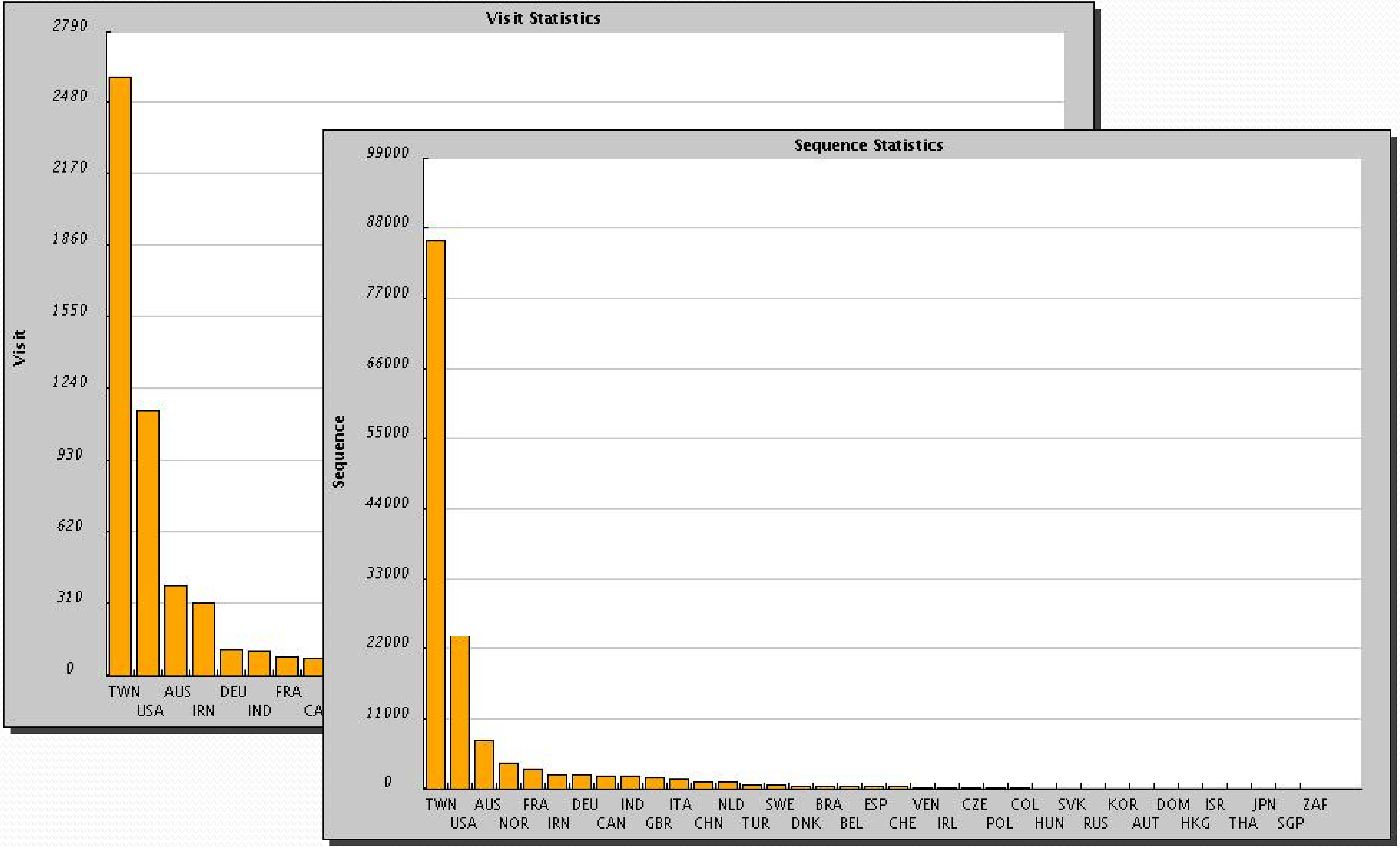
Ching-Chun Chang,[1] Hsien-Chia Lin,*[1] I-Pin Lin,† Teh-Yuan Chow,‡[2] Hong-Hwa Chen,* Wen-Huei Chen,§ Chia-Hsiung Cheng,‡ Chung-Yen Lin,|| Shu-Mei Liu,‡ Chien-Chang Chang,¶ and Shu-Miaw Chaw¶*

*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

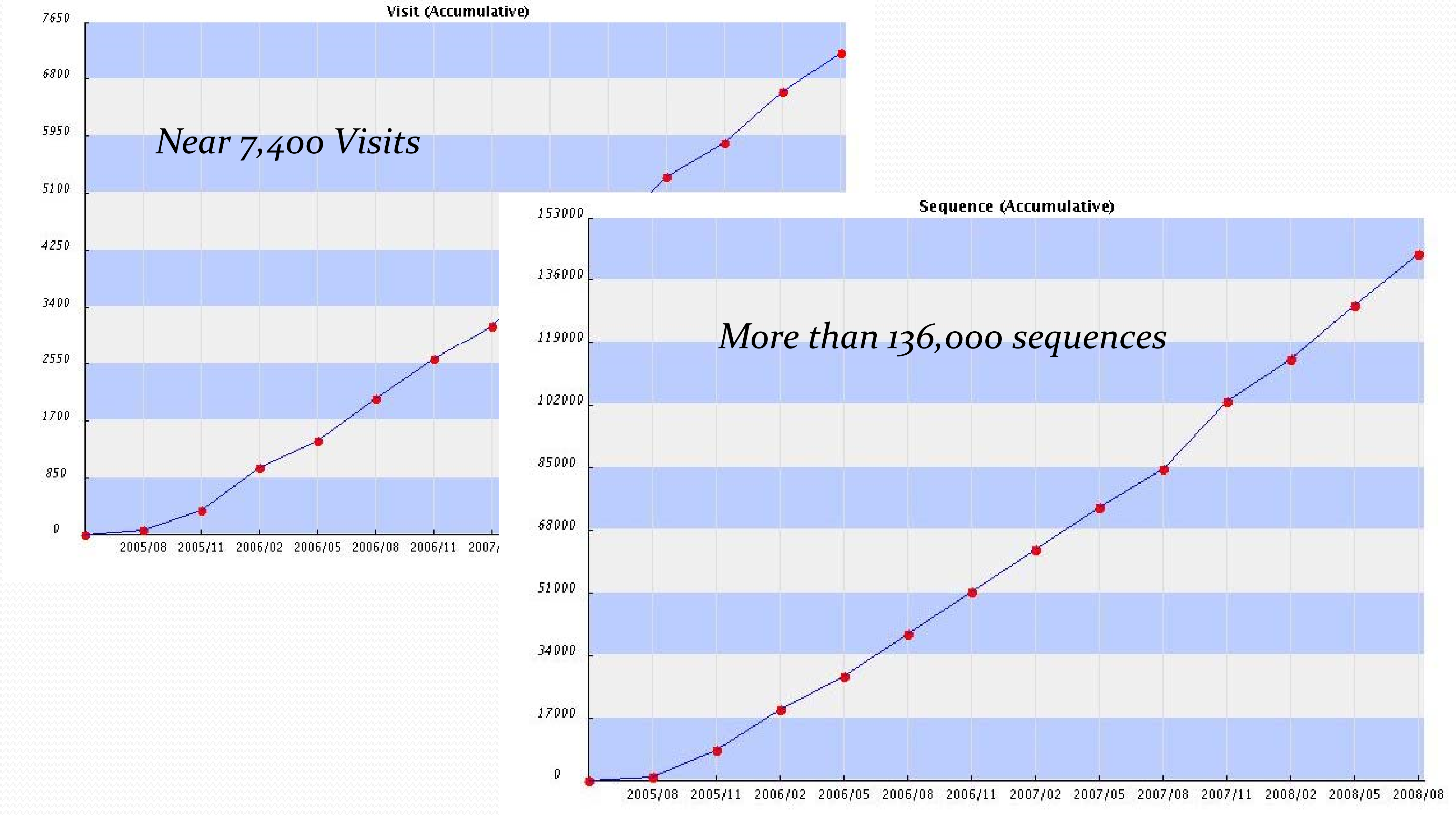

LSC
Taiwan Moth Orchid
(*Phalaenopsis aphrodite*)
Chloroplast DNA
148,964 bp
J_LB
J_LA
IR_B
IR_A
J_SB
J_SA
SSC
Genes for photosystem I, photosystem II and cytochrome b6/f complex
Genes for subunits of ATP synthase complex
Genes for RuBisCO large subunit
Genes for subunits of NADH dehydrogenase
Gene for 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



Service Usage of POWER *from 2005 July.*



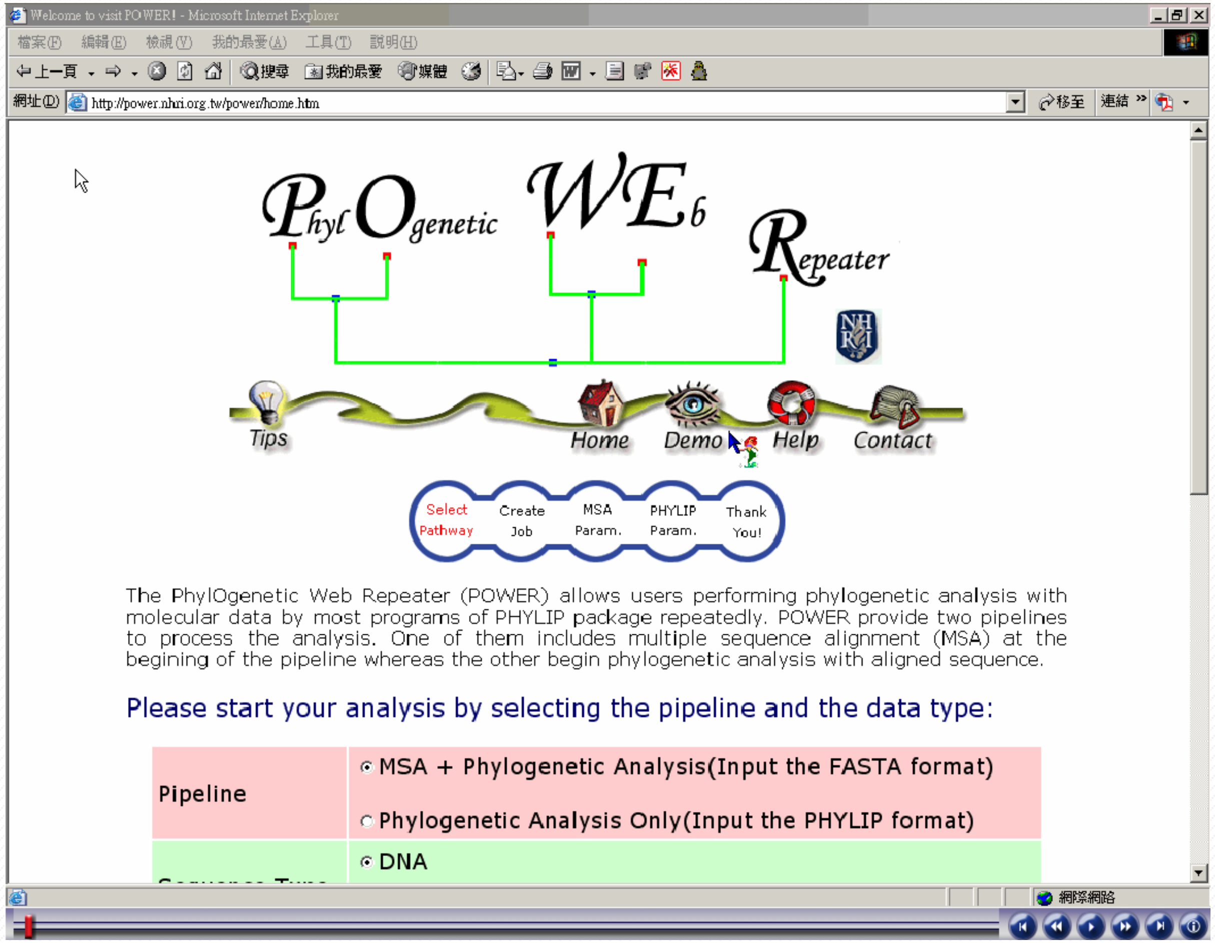
Service Usage of POWER *from 2005 July.*



Near 7,400 Visits

More than 136,000 sequences

Automatic On-Line Demonstration

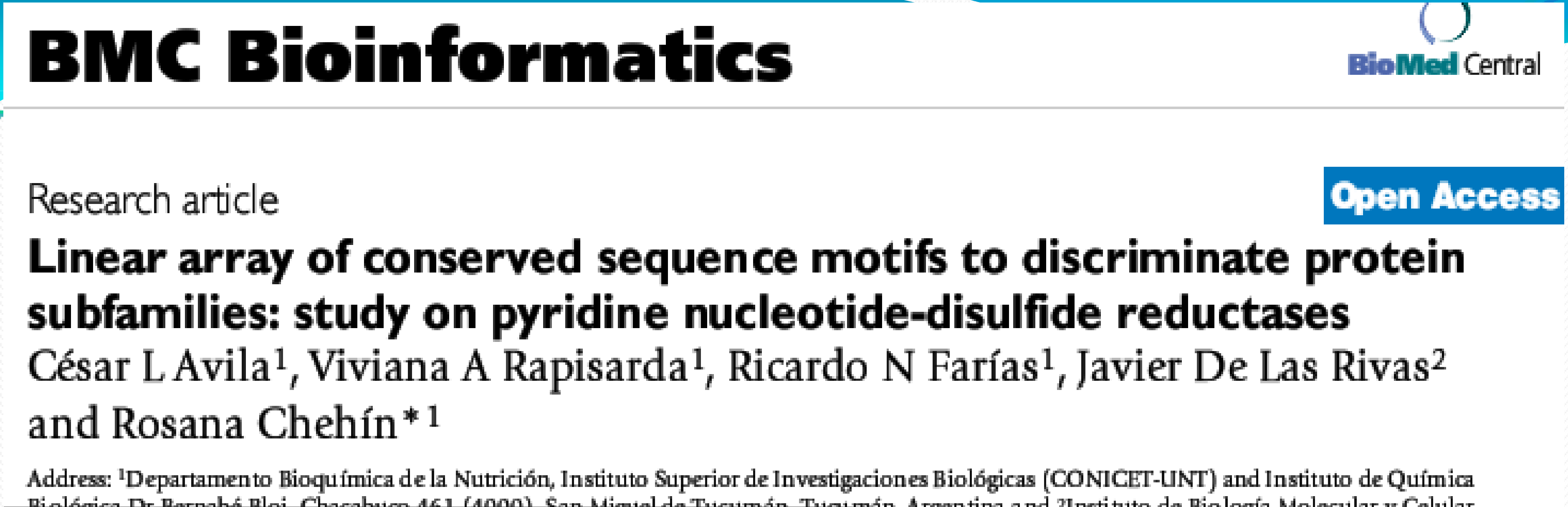

Welcome to visit POWER! - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://power.nhri.org.tw/power/home.htm
移至 連結
PhylOgenetic WEb Repeater
Tips
Home
Demo
Help
Contact
Select Pathway
Create Job
MSA Param.
PHYLIP Param.
Thank You!
網際網路

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

http://www.nhri.org.tw/nhri_org/bs/biostat/power.swf



BMC Bioinformatics

BioMed Central

Research article

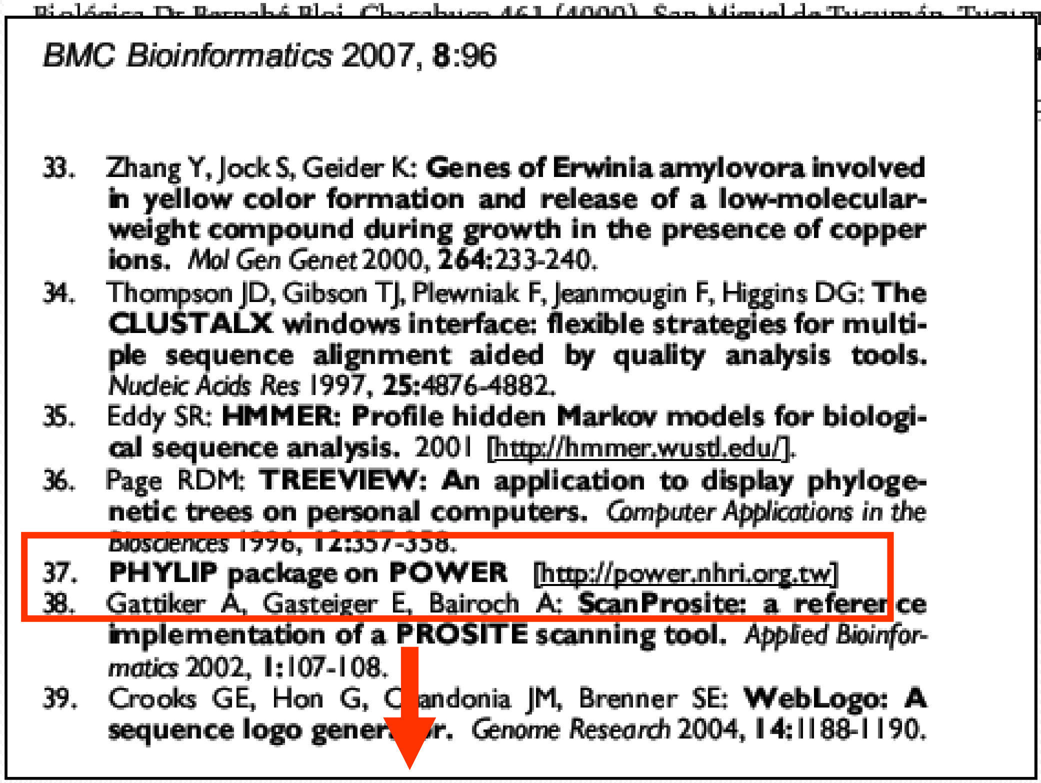
Open Access

Linear array of conserved sequence motifs to discriminate protein subfamilies: study on pyridine nucleotide-disulfide reductases

César L Avila[1], Viviana A Rapisarda[1], Ricardo N Farías[1], Javier De Las Rivas[2] and Rosana Chehín*[1]

Address: [1]Departamento Bioquímica de la Nutrición, Instituto Superior de Investigaciones Biológicas (CONICET-UNT) and Instituto de Química Biológica Dr Bernabé Bloj, Chacabuco 461 (4000), San Miguel de Tucumán, Tucumán, Argentina and [2]Instituto de Biología Molecular y Celular
, Spain

qf.unt.edu.ar; Ricardo N Farías - rfarias@conicet.gov.ar;



BMC Bioinformatics 2007, **8**:96

33. Zhang Y, Jock S, Geider K: **Genes of Erwinia amylovora involved in yellow color formation and release of a low-molecular-weight compound during growth in the presence of copper ions.** *Mol Gen Genet* 2000, **264:**233-240.

34. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25:**4876-4882.

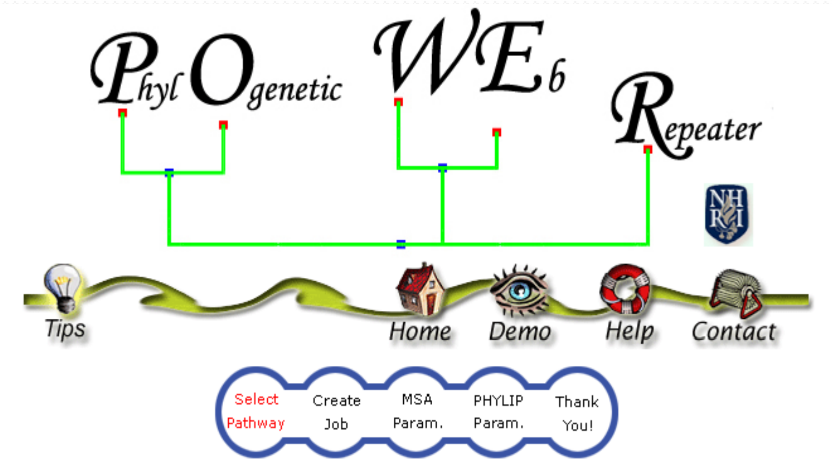
35. Eddy SR: **HMMER: Profile hidden Markov models for biological sequence analysis.** 2001 [http://hmmer.wustl.edu/].

36. Page RDM: **TREEVIEW: An application to display phylogenetic trees on personal computers.** *Computer Applications in the Biosciences* 1996, **12:**357-358.

37. **PHYLIP package on POWER** [http://power.nhri.org.tw]

38. Gattiker A, Gasteiger E, Bairoch A: **ScanProsite: a reference implementation of a PROSITE scanning tool.** *Applied Bioinformatics* 2002, **1:**107-108.

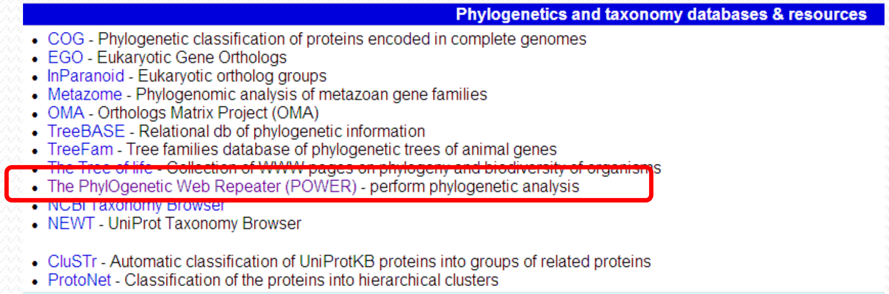
39. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: A sequence logo generator.** *Genome Research* 2004, **14:**1188-1190.



PHYLIP package on POWER [http://power.nhri.org.tw]

POWER Listed in Bioinfo Portals

- 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)

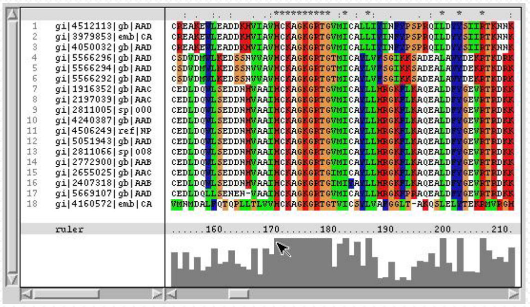


Phylogenetics and taxonomy databases & resources

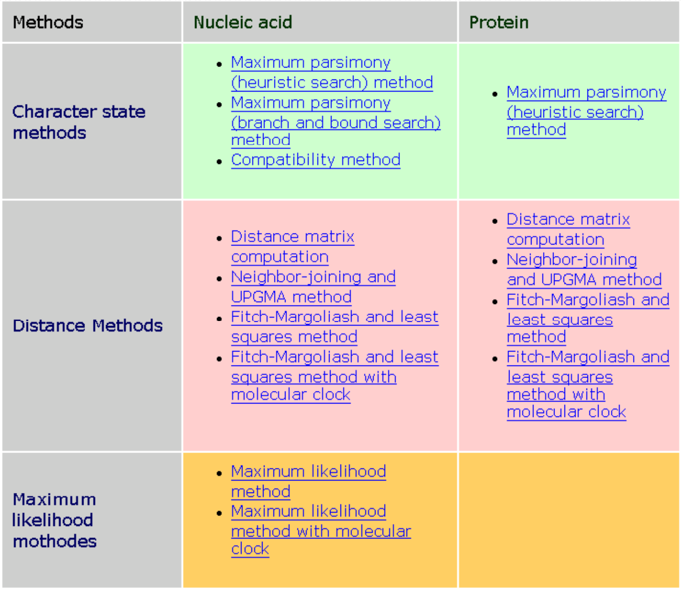
- COG - Phylogenetic classification of proteins encoded in complete genomes
- EGO - Eukaryotic Gene Orthologs
- InParanoid - Eukaryotic ortholog groups
- Metazome - Phylogenomic analysis of metazoan gene families
- OMA - Orthologs Matrix Project (OMA)
- TreeBASE - Relational db of phylogenetic information
- 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

General Pipeline for Phylogenetic Analysis

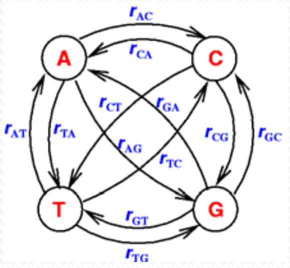


Multiple Sequence Alignment

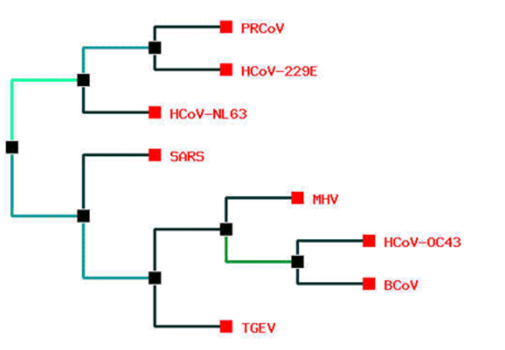


Methods	Nucleic acid	Protein
Character state methods	• Maximum parsimony (heuristic search) method • Maximum parsimony (branch and bound search) method • Compatibility method	• Maximum parsimony (heuristic search) method
Distance Methods	• Distance matrix computation • Neighbor-joining and UPGMA method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

Selection of inference Methods



Bootstrap
Substitution Model
Tree Construction

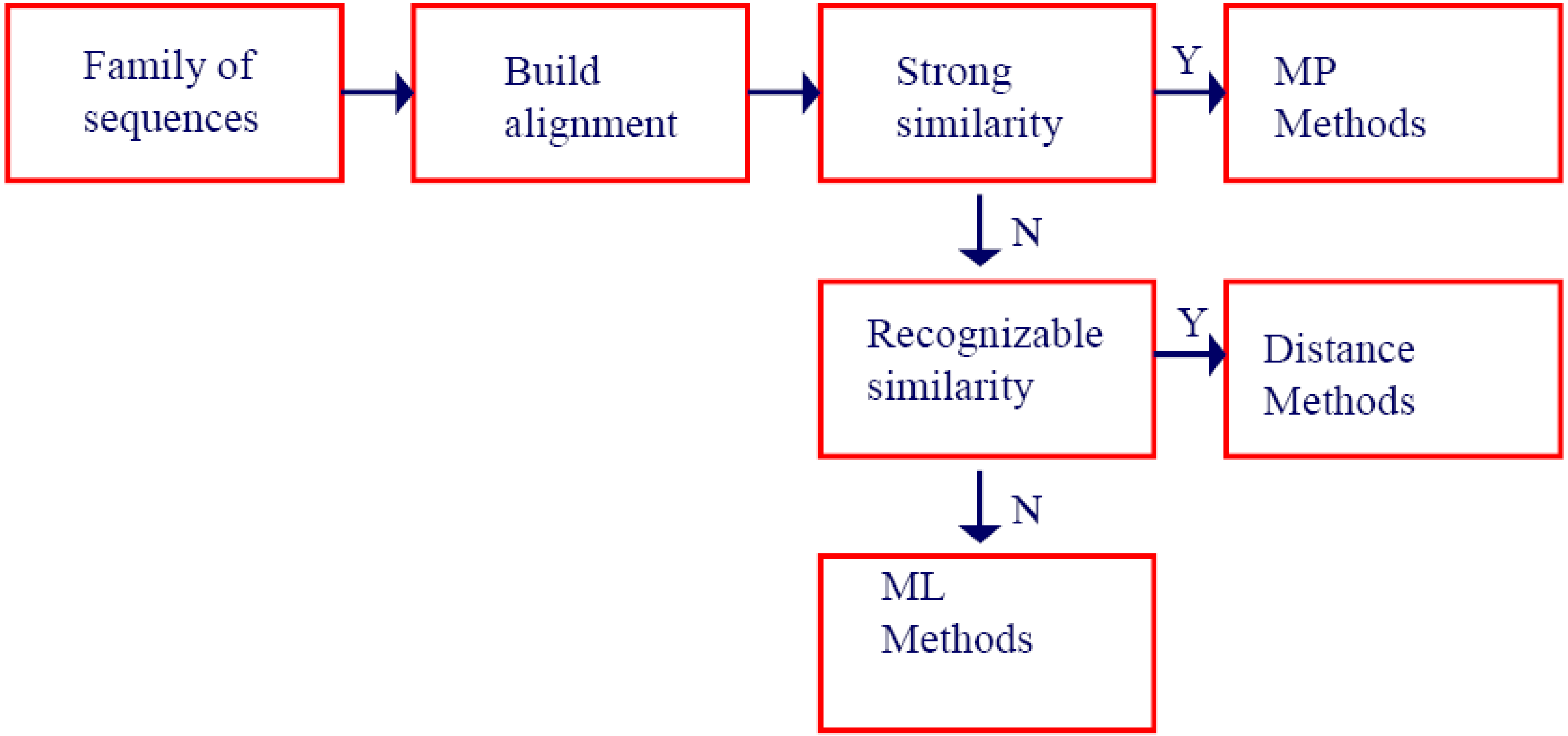


Evaluate phylogenetic tree

Phylogenetic Analysis

- Character state method
  - Maximum parsimony
- Distance method
  - Neighbor-joining and UPGMA method
  - Fitch-Margoliash method
- Maximum likelihood methods
  - determinate evolution model first, then construct system trees

Flowchart of Analysis



(Mount, *Bioinformatics*)

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 ɵ, and the tree τ, including topology and branch lengths)

$$L(\tau, \theta\ )=\text{Prob}(\text{Data}|\ \tau, \theta\ )$$

Or

Prob(Aligned Sequences| tree, model of evolution)

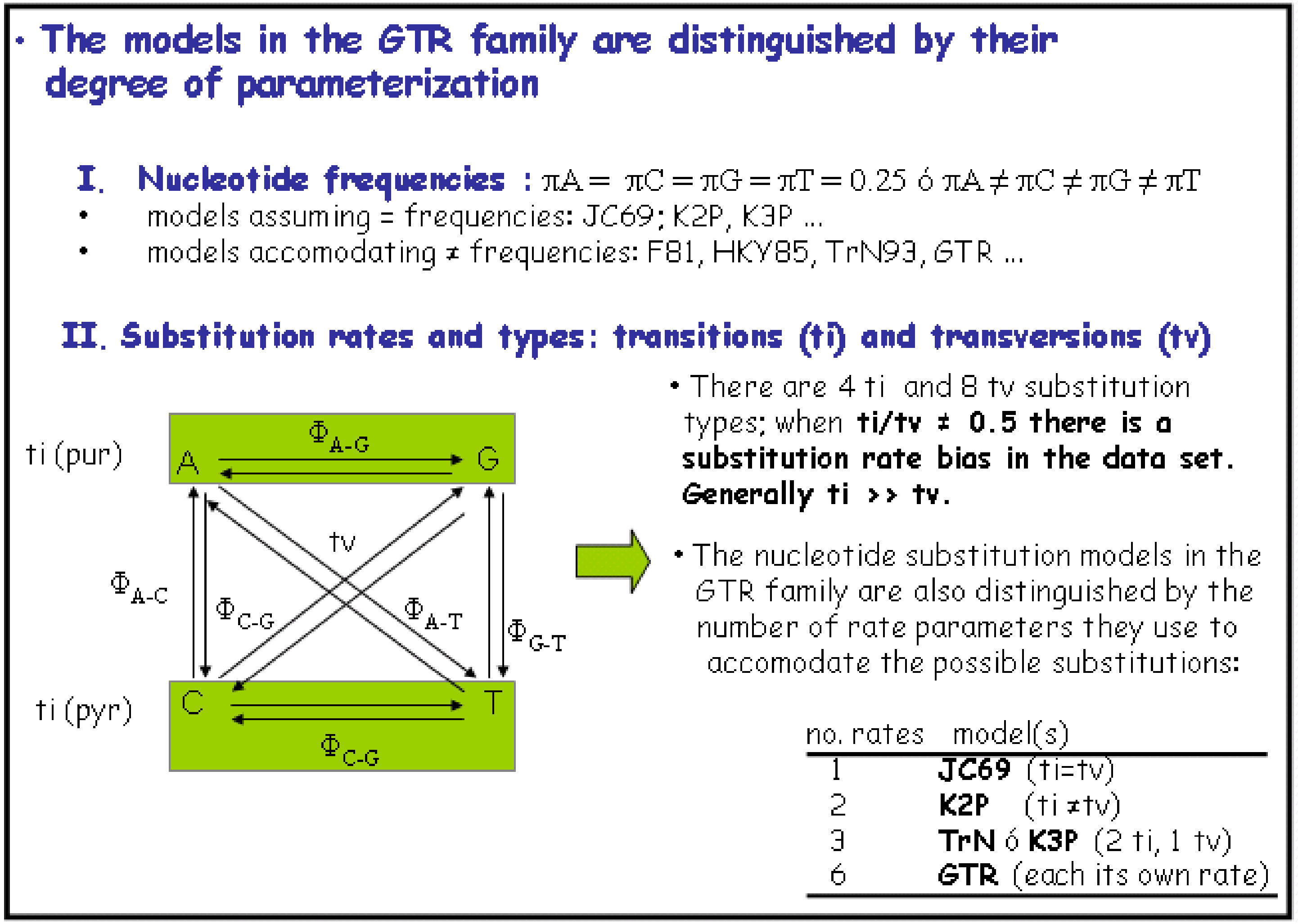
Maximum Likelihood Estimates (MLE)

- The maximum likelihood estimates (MLE) of τ, ɵ are those making the LH function as large as possible

$$\tau, \theta = \max L(\tau, \theta)$$

- Hence, what we usually call the likelihood of the tree is **not the likelihood of the tree**, but **the probability of the data given that the tree is the true tree**.

Basic Substitution Model



- **The models in the GTR family are distinguished by their degree of parameterization**

I. Nucleotide frequencies : $\pi A = \pi C = \pi G = \pi T = 0.25$ ó $\pi A \neq \pi C \neq \pi G \neq \pi T$

- models assuming = frequencies: JC69; K2P, K3P ...
- models accomodating ≠ frequencies: F81, HKY85, TrN93, GTR ...

II. Substitution rates and types: transitions (ti) and transversions (tv)

ti (pur)
$\Phi_{A\text{-}G}$
A
G
tv
$\Phi_{A\text{-}C}$
$\Phi_{C\text{-}G}$
$\Phi_{A\text{-}T}$
$\Phi_{G\text{-}T}$
ti (pyr)
C
T
$\Phi_{C\text{-}G}$

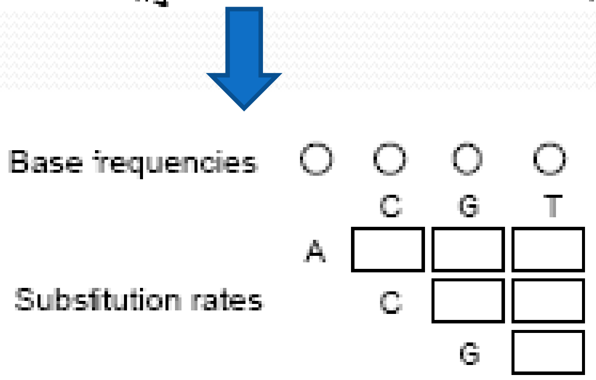
- There are 4 ti and 8 tv substitution types; when **ti/tv ≠ 0.5 there is a substitution rate bias in the data set. Generally ti >> tv.**

- The nucleotide substitution models in the GTR family are also distinguished by the number of rate parameters they use to accomodate the possible substitutions:

no. rates	model(s)
1	**JC69** (ti=tv)
2	**K2P** (ti ≠tv)
3	**TrN** ó **K3P** (2 ti, 1 tv)
6	**GTR** (each its own rate)

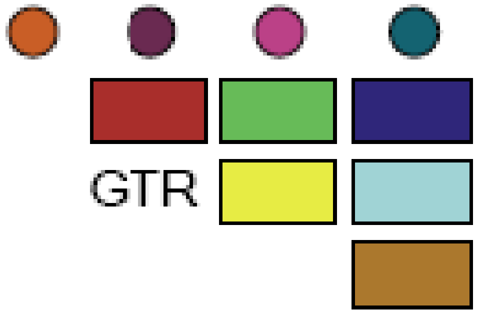
Illustration of DNA substitution Model

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



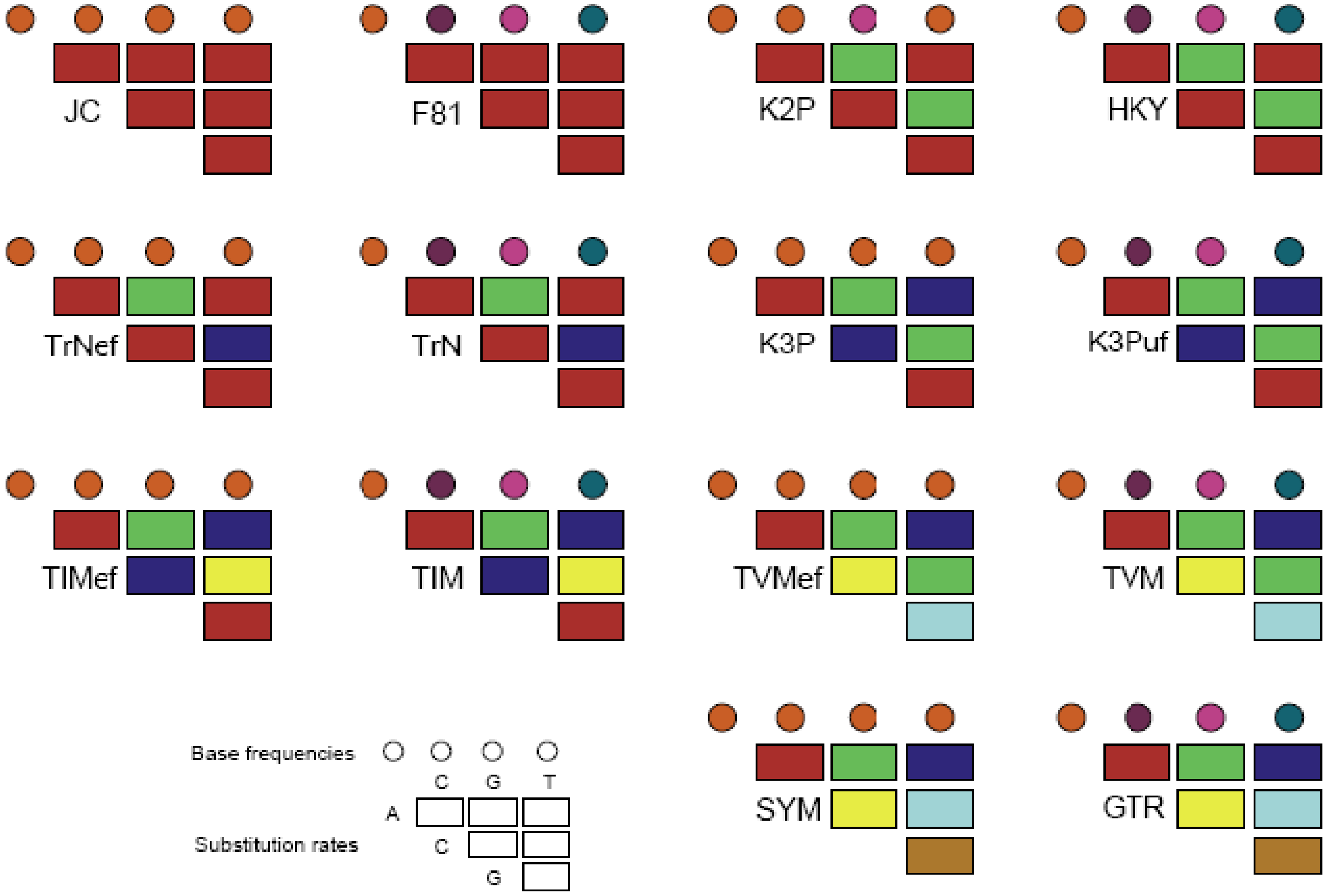
Base frequencies ○ ○ ○ ○

Substitution rates



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



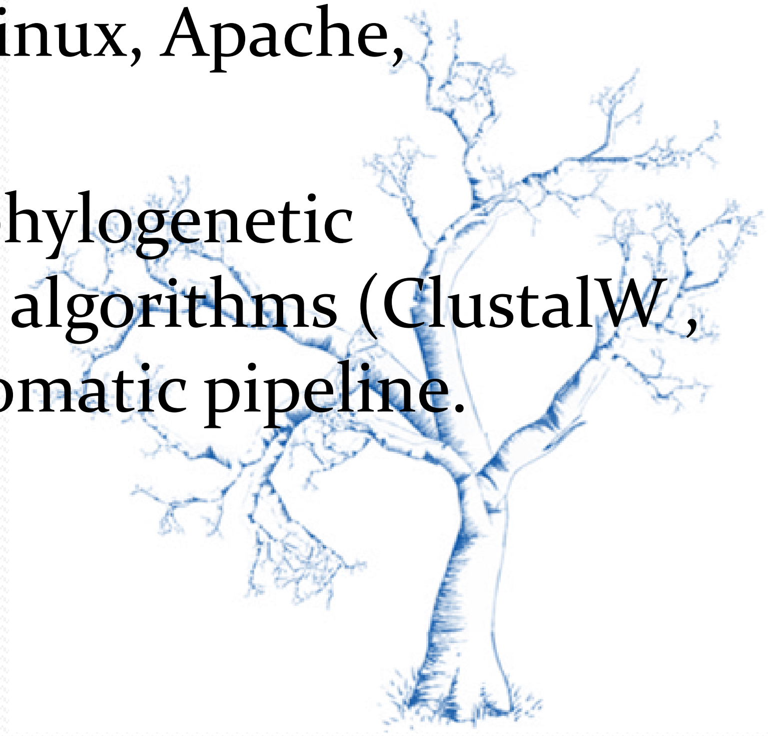
Background

- Model fitting in phylogenetics has been suggested for many years, yet many authors still arbitrarily choose their models, often using the default models implemented in standard computer programs for phylogenetic estimation.
- Here, we want to show the way that a best-fit model can be readily identified. Consequently, given the relevance of models, model fitting should be routine in any phylogenetic analysis that uses models of evolution.

http://www.ncbi.nlm.nih.gov/pubmed/12116655?dopt=Abstract

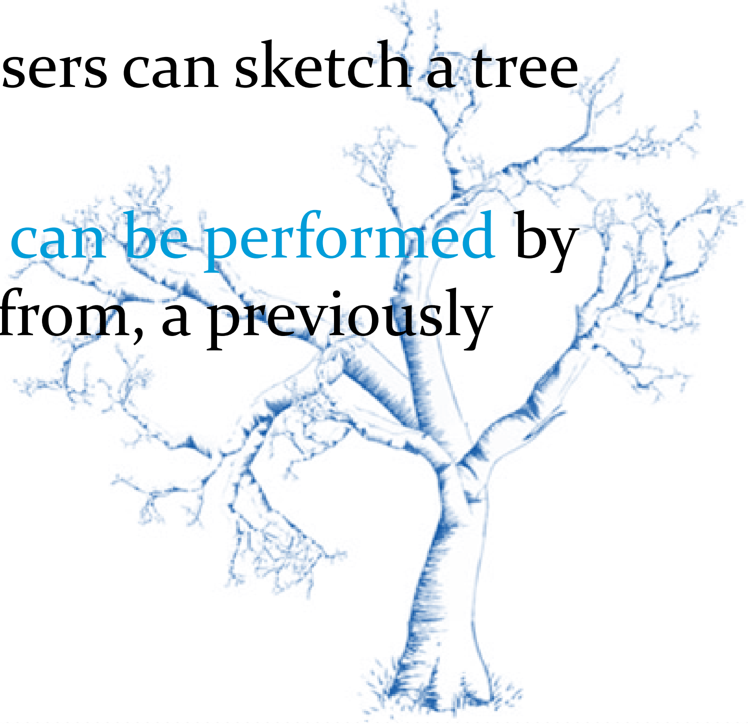
Motivation I

- Provide a seamless way to conduct the complex phylogenetic analysis for Biologists
- An integrated and user-optimized framework for biomolecular phylogenetic analysis
- PALM uses an open-source LAPP (Linux, Apache, PostgreSql, PHP) structure and
- PALM infers genetic distances and phylogenetic relationships using well-established algorithms (ClustalW , PhyML, ProtTest, Modeltest) in automatic pipeline.



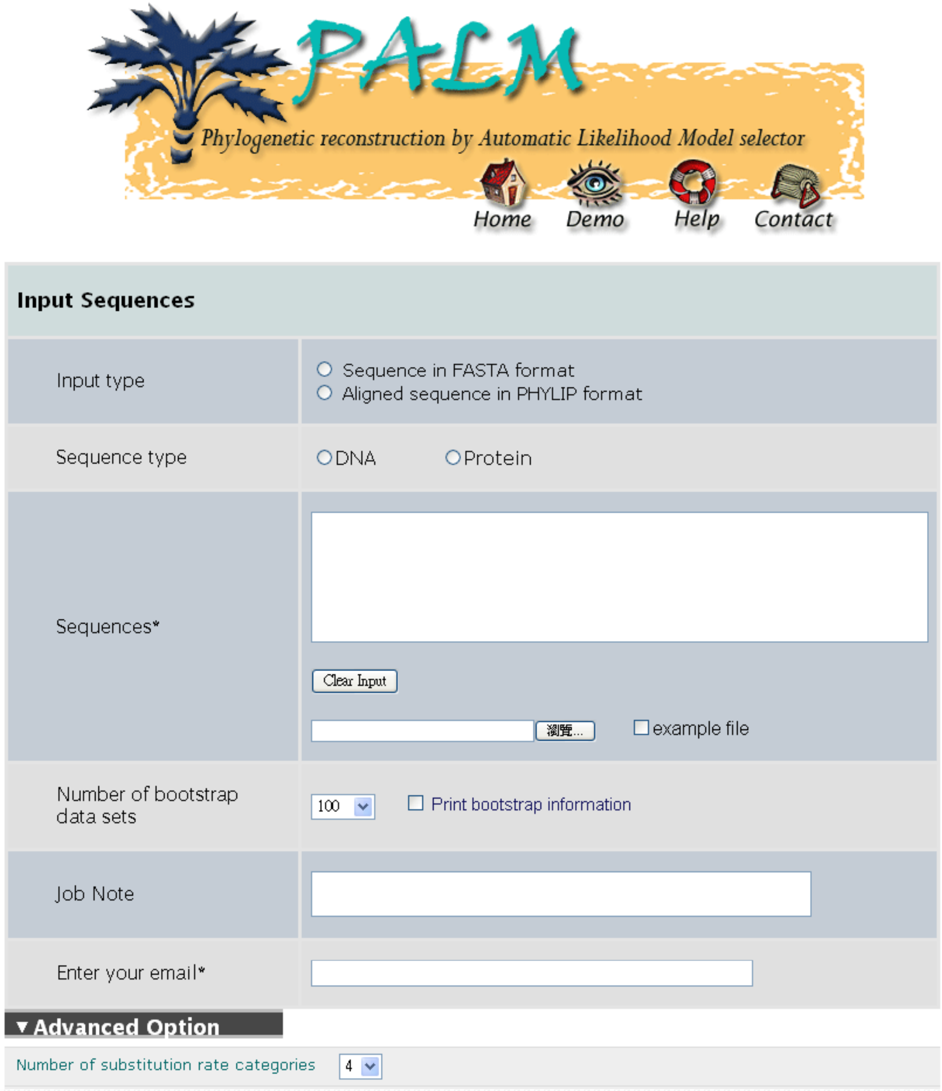
Motivation II

- Model can be selected by following methods including hierarchical likelihood ratio tests (hLRTs), Akaike information criterion (AIC), and Bayesian information criterion (BIC)
- PALM can help user to construct the tree with bootstrap based on best substitution model chosen by maximum likelihood.
- 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



Component Programs of PALM

- PhyML 3.0
- ModelTest 3.7
- ProtTest 1.4
- ClustalW 2.0.3
- Seqret (EMBOSS)



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*	

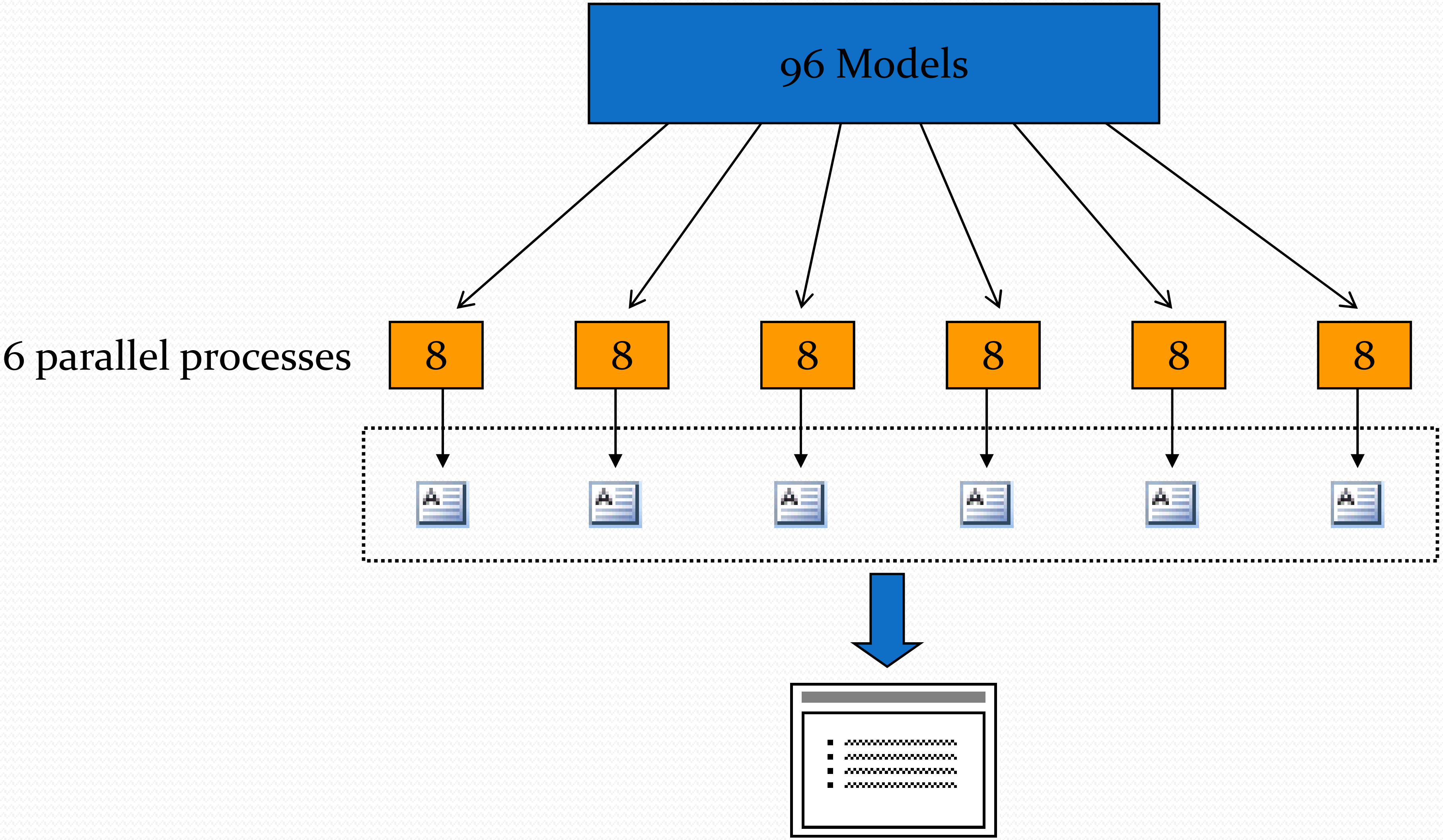
▼Advanced Option

Number of substitution rate categories 4

Models Used in PALM

- For DNA (24 models)
  - JC69, K80, F81, HKY, TrN, GTR
  - +I, +G
- For Protein (96 models), **Time consuming**
  - JTT, MtREV, MtMam, MtArt, Dayhoff, WAG, RtREV, CpREV, Blosum62, VT, HIVb, HIVw
  - +I, +G, +F

PalmMonitor for Protein Models



Decreasing Time by PALMmonitor

- According the algorithm used in PALM, some models will take a lot of time to calculate the value of maximum likelihood.

JTT MtREV	3h:00:50
MtMam MtArt	3h:29:04
Dayhoff WAG	2h:50:16
RtREV CpREV	2h:50:19
Blosum62 VT	2h:49:17
HIVb HIVw	2h:56:38

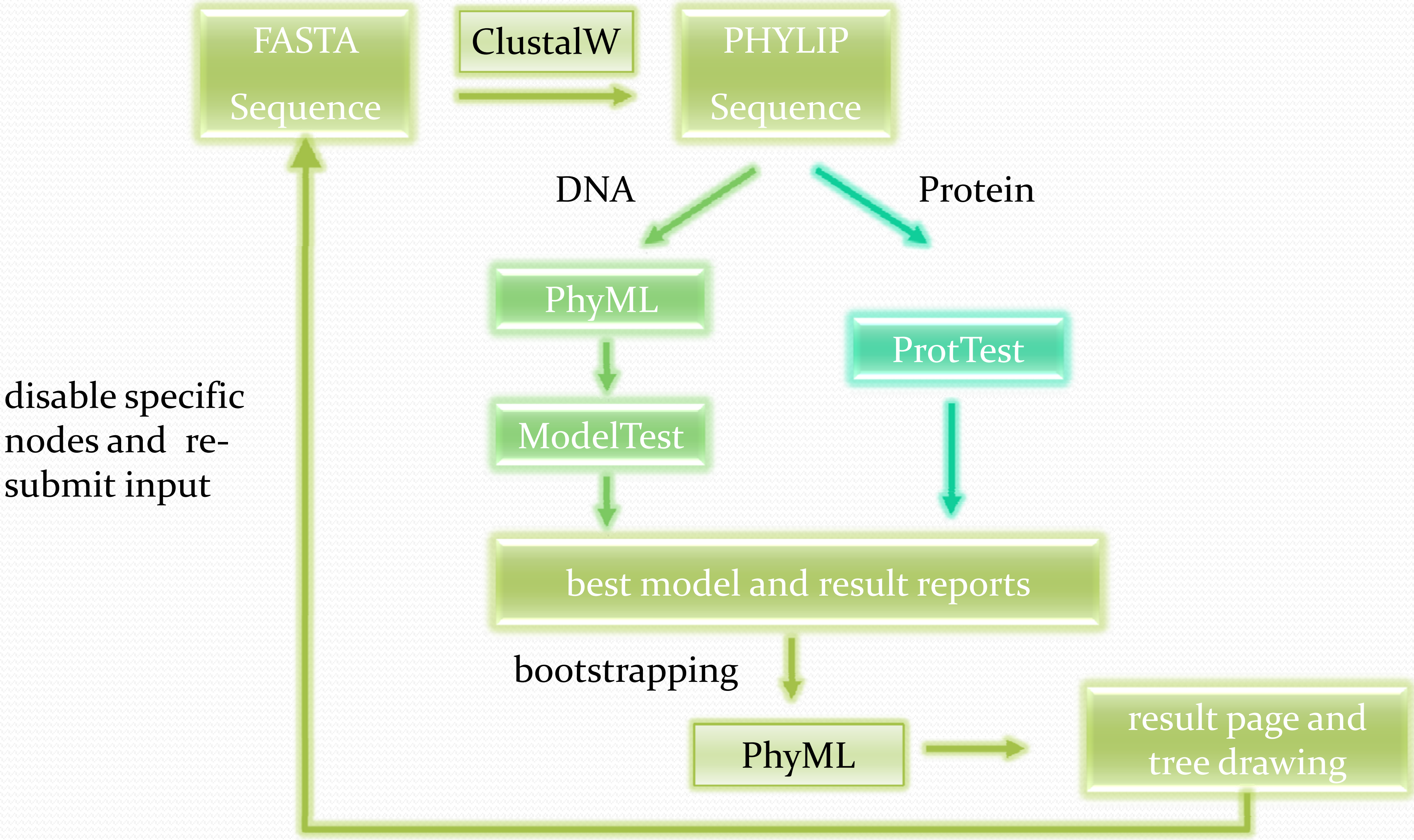
- All Models 7h:32:10

Source: 25 sequences with 5000 residues for each

Input and Output of PALM

- Input format (Protein and DNA)
  - Fasta format
  - Phylip format: Aligned Sequences
  - User tree (if submitted and valid)
- Output
  - Tree topology by php and GD library
  - Tree file in Newick format
  - Aligned Sequence in phylip format
  - Best model selector by PALM

Flowchart of PALM



FASTA Sequence

ClustalW

PHYLIP Sequence

DNA

Protein

PhyML

ModelTest

ProtTest

best model and result reports

bootstrapping

PhyML

result page and tree drawing

disable specific nodes and re-submit input

Result of PALM

PALM

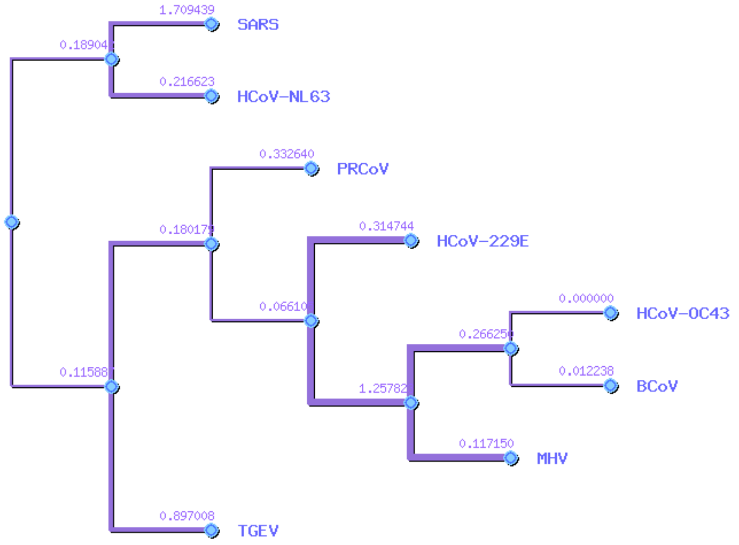
Phylogenetic reconstruction by Automatic Likelihood Model selector

Home Demo Help Contact

PALM Result

Input parameters

Job ID	20080525234728289	**Number of Substitution Rate Category**	4
Sequence Type	DNA	**Model Selection Criterion**	AICc
Number of Bootstrap	100	**Optimization of Tree Topology**	Yes
Job Note	test in 100 BS	**Optimization of Branch Length**	Yes
Starting Tree	BIOJ		



Result Information

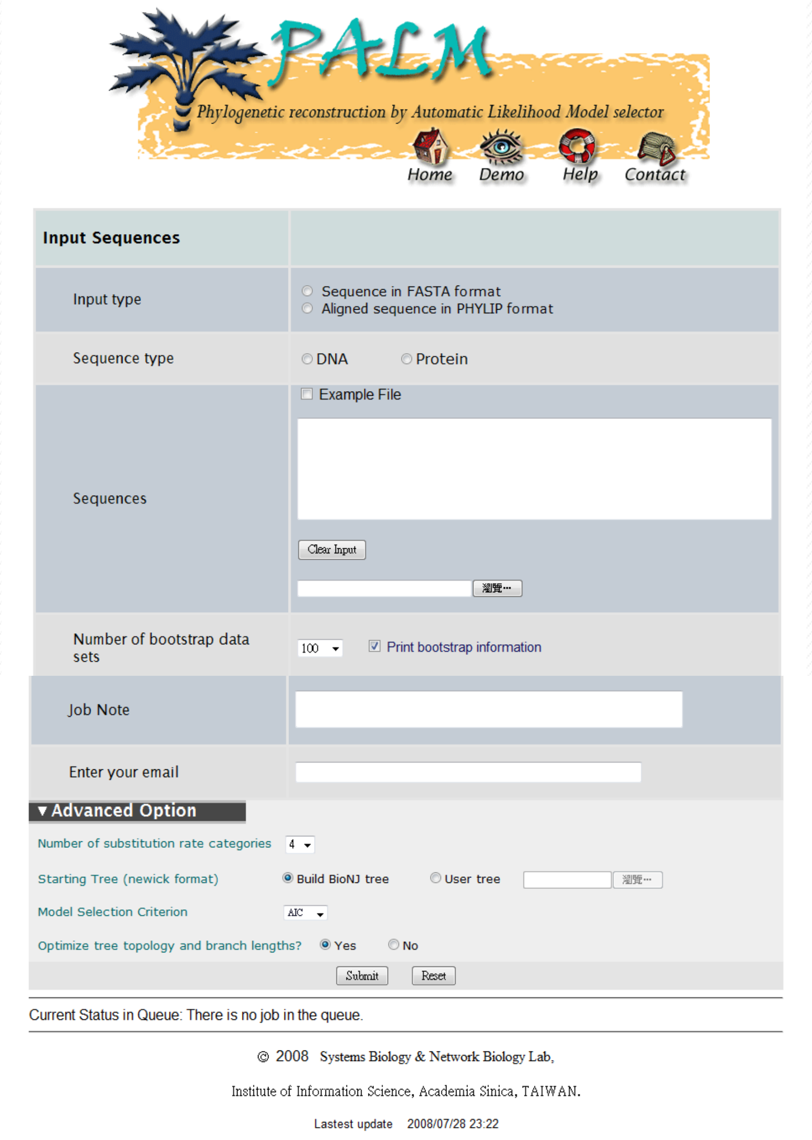
Best Model Selected	GTR+G
Model Selection Criterion	AICc
-lnL	1903.3464
Number of Estimated Parameters (K)	9

Model	-lnL	K	AICc	delta	weight	cumWeight
GTR+G	1903.3464	9	3825.0603	3819.0361	0.00e+00	1.0000
GTR	1905.8035	8	3827.9001	3821.8760	0.00e+00	1.0000
GTR+I+G	1904.1582	10	3828.7664	3822.7422	0.00e+00	1.0000
GTR+I	1905.8112	9	3829.9897	3823.9656	0.00e+00	1.0000
TrN+G	1910.6451	6	3833.4607	3827.4365	0.00e+00	1.0000
HKY	1912.7156	4	3833.5120	3827.4878	0.00e+00	1.0000
TrN	1911.7296	5	3833.5806	3827.5564	0.00e+00	1.0000
HKY+G	1911.9691	5	3834.0596	3828.0354	0.00e+00	1.0000
TrN+I+G	1910.6479	7	3835.5234	3829.4993	0.00e+00	1.0000
HKY+I	1912.7211	5	3835.5635	3829.5393	0.00e+00	1.0000
TrN+I	1911.7354	6	3835.6411	3829.6169	0.00e+00	1.0000
HKY+I+G	1911.9722	6	3836.1147	3830.0906	0.00e+00	1.0000
F81+G	1941.3434	4	3890.7676	3884.7434	0.00e+00	1.0000
K80	1945.1681	1	3892.3442	3886.3201	0.00e+00	1.0000
F81+I+G	1941.3442	5	3892.8098	3886.7856	0.00e+00	1.0000
F81	1943.7166	3	3893.4814	3887.4573	0.00e+00	1.0000
K80+G	1944.9779	2	3893.9800	3887.9558	0.00e+00	1.0000

Download Area

Original File	20080525234728289
Phylip File	20080525234728289.phy
Phylogenetic Tree (Newick)	tree20080525234728289.txt
Statistic data	20080525234728289_phyml_stat.txt
Modelselection Information	Modeltest20080525234728289.out
Bootstrap Tree	20080525234728289_phyml_boot_trees.txt
Bootstrap Statistic data	20080525234728289_phyml_boot_stats.txt

Demonstration of PALM



PALM

Phylogenetic reconstruction by Automatic Likelihood Model selector

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

▾ Advanced Option

Number of substitution rate categories	4	
Starting Tree (newick format)	● Build BioNJ tree	○ User tree 瀏覽...
Model Selection Criterion	AIC	
Optimize tree topology and branch lengths?	● Yes	○ No

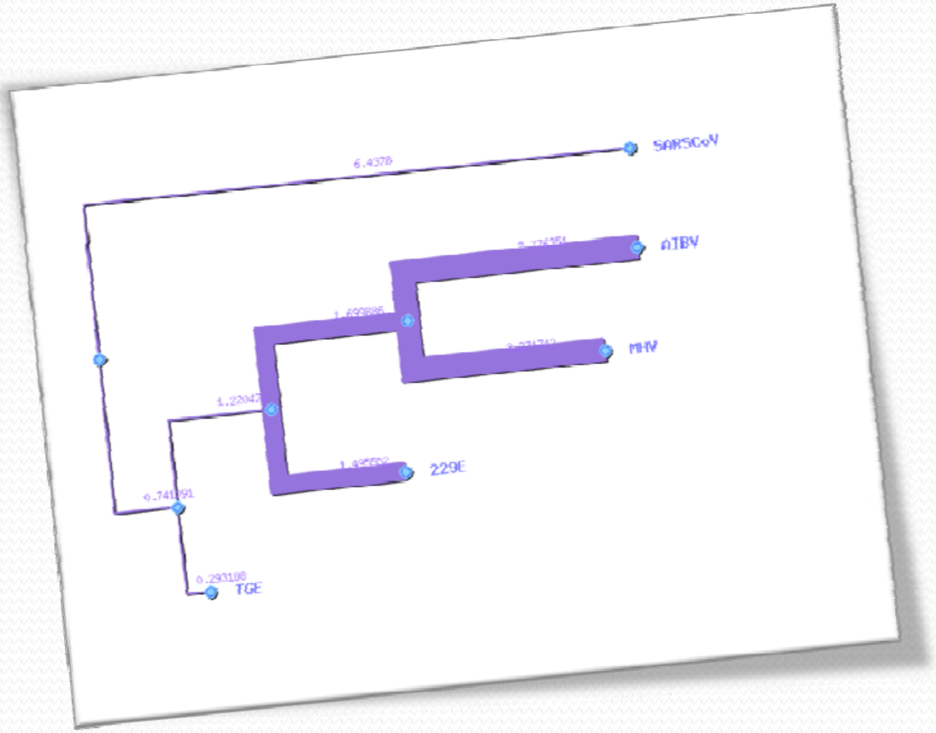
Submit Reset

Current Status in Queue: There is no job in the queue.

© 2008 Systems Biology & Network Biology Lab,

Institute of Information Science, Academia Sinica, TAIWAN.

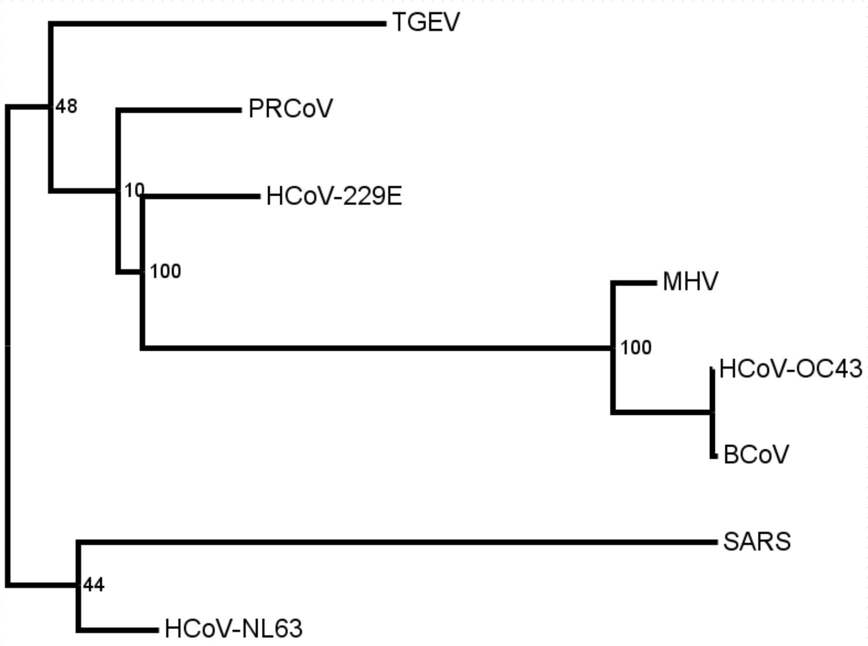
Lastest update 2008/07/28 23:22



Access : http://palm.iis.sinica.edu.tw

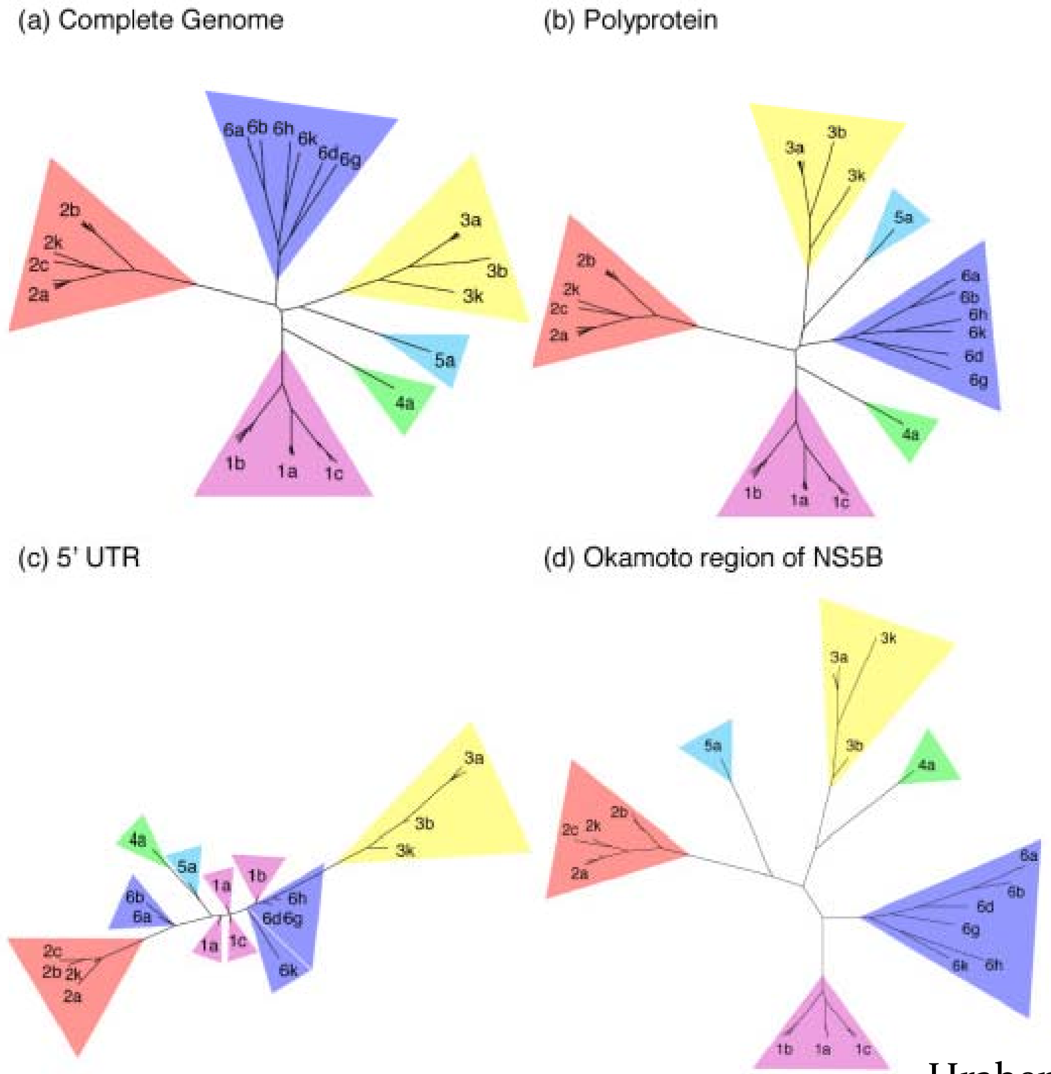
Bootstrap (BS) Analysis

- Bootstrap analysis is the most often used 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 due to the original bias connot 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

- Integrate more substitution models into PALM
- Improve and optimize the performance of whole pipeline
- MrBayes will be implemented into this system for Bayesian inference.

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

Acknowledgement

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

Daniel, Sheng-Yao Su

Tengi, Huang

Pao-Han Kuo

Chen-Ren Lo

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

Protein Network

✓*hp*-DPI

✓*fly*DPI

✓Reconstruction of Human protein network

✓Topological analysis by Hubba

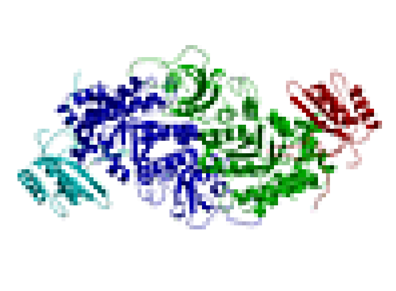
Helicobacter Pylori
Database of Protein Interactomes
Combined with Experimental and Inferring Interactions
Help
Contact

Motivations

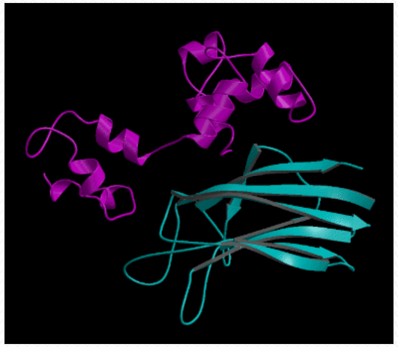
- Combine accumulated fragmentary information (experimental interactions) into a systems-level picture (embrace experimental and putative interactions) with spatiotemporal scenarios
- Construct entire network including those interactions can’t be done due to experiment limitation (ie. Toxic and membrane proteins can not be tested in Y2H).
- Understand host and pathogen networks, how they merge during infection
- Provide a multilayered and integrated view to control diseases ranging pathogenic infection to cancer

Deciphering Protein into Domains

- Using the protein-protein interaction (PPI) data set to infer domain-domain interaction (DDI) for specific organism. Then using the predicted DDI set can infer other probable PPI set
- Deciphering the domain interaction will allows us to discover novel interactions between proteins that contain domains with known binding partner.



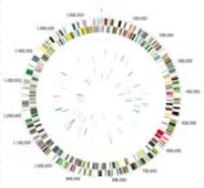
Protein Interactions

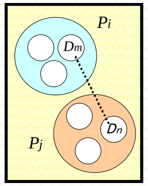


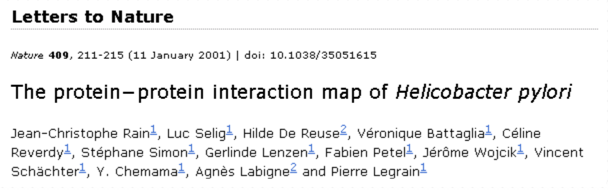
Domain Interactions

Previous Work I:

Helicobacter pylori- Database of Protein Interactome







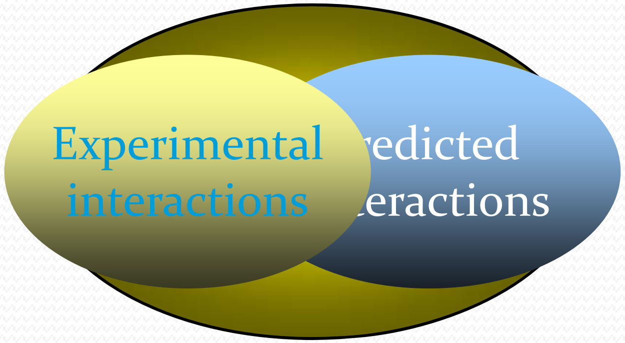
Letters to Nature

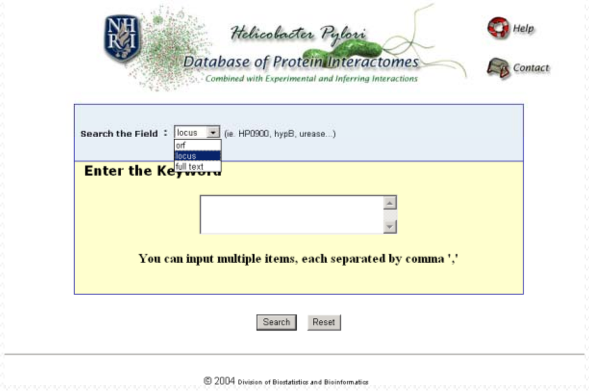
Nature **409**, 211-215 (11 January 2001) | doi: 10.1038/35051615

The protein–protein interaction map of *Helicobacter pylori*

Jean-Christophe Rain[1], Luc Selig[1], Hilde De Reuse[2], Véronique Battaglia[1], Céline Reverdy[1], Stéphane Simon[1], Gerlinde Lenzen[1], Fabien Petel[1], Jérôme Wojcik[1], Vincent Schächter[1], Y. Chemama[1], Agnès Labigne[2] and Pierre Legrain[1]

Over 1,200 interactions were identified between H. pylori (strain 26695) proteins, connecting 46.6% of the proteome.




Helicobacter Pylori
Database of Protein Interactomes
Combined with Experimental and Inferring Interactions
Help
Contact
locus
orf
locus
full text
Search
Reset

Search the Field : locus (ie. HP0900, hypB, urease...)

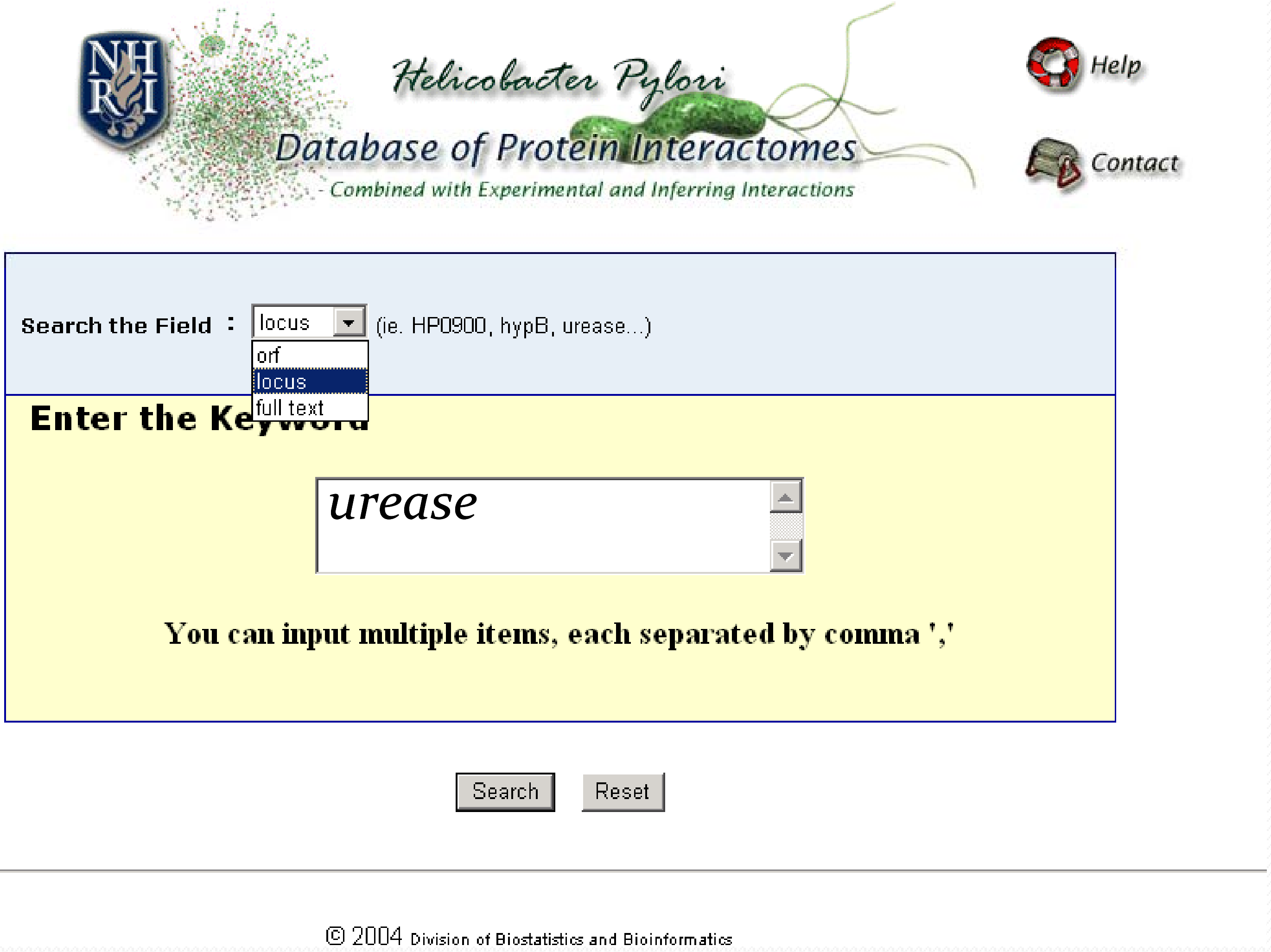
Enter the Keyword

You can input multiple items, each separated by comma ','

© 2004 Division of Biostatistics and Bioinformatics

The network of whole proteome

Previous Work, hp-DPI



NHRI
Helicobacter Pylori
Database of Protein Interactomes
- Combined with Experimental and Inferring Interactions
Help
Contact
orf
locus
full text
Search
Reset

Search the Field : locus (ie. HP0900, hypB, urease...)

Enter the Keyword

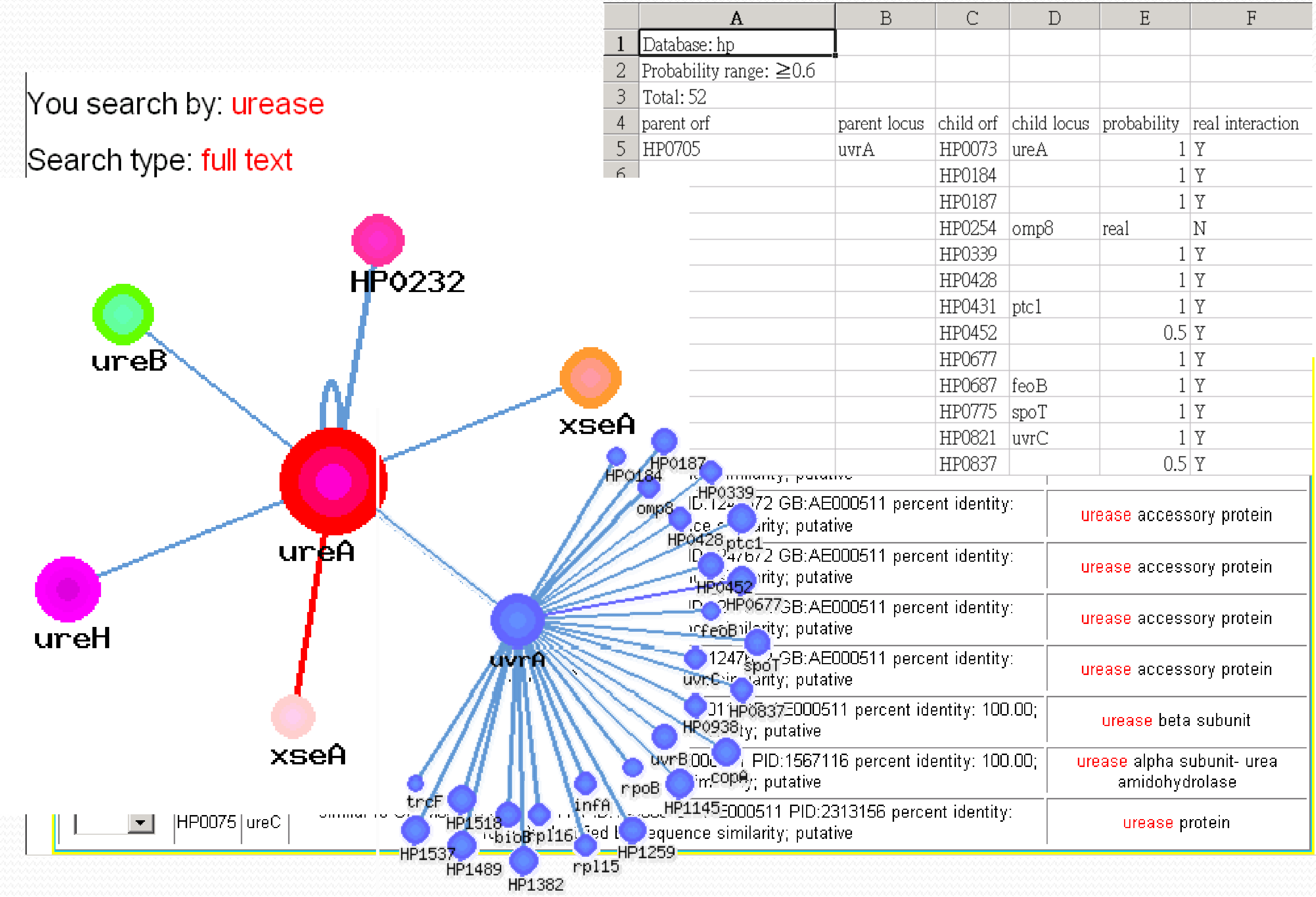
urease

You can input multiple items, each separated by comma ','

© 2004 Division of Biostatistics and Bioinformatics

This website can be accessed at http://dpi.nhri.org.tw/hp/

Search Result of hp-DPI



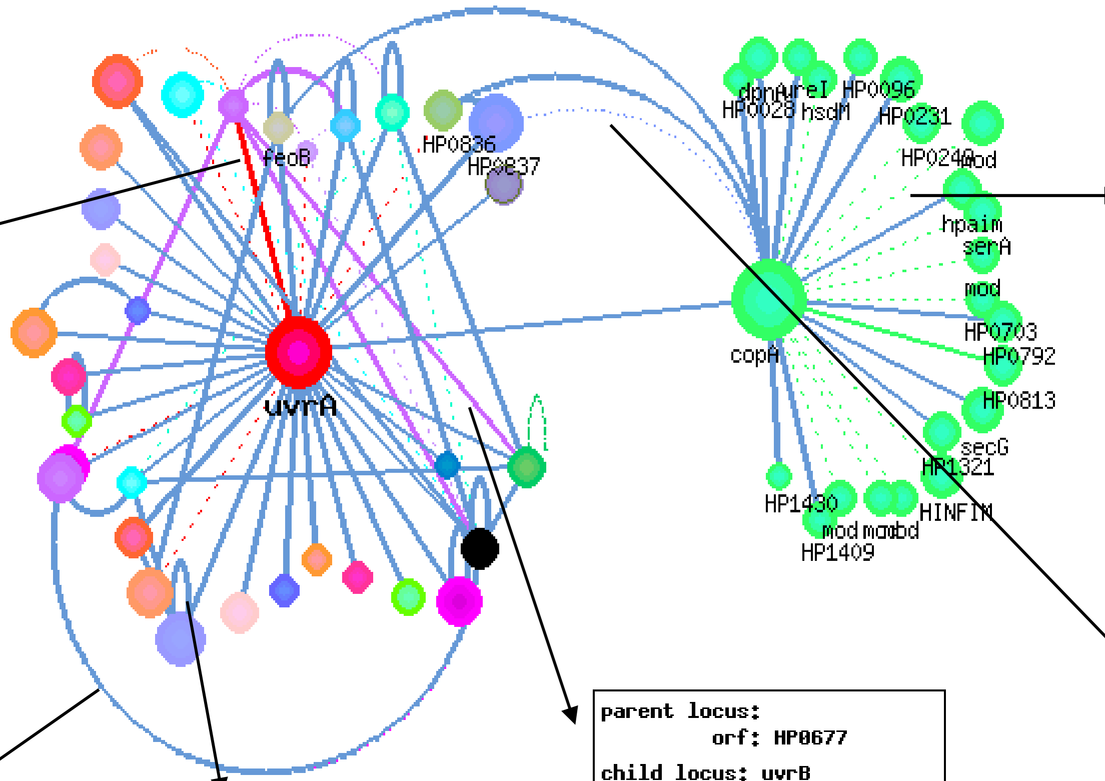
You search by: urease

Search type: full text

	A	B	C	D	E	F
1	Database: hp					
2	Probability range: ≧0.6					
3	Total: 52					
4	parent orf	parent locus	child orf	child locus	probability	real interaction
5	HP0705	uvrA	HP0073	ureA	1	Y
6			HP0184		1	Y
			HP0187		1	Y
			HP0254	omp8	real	N
			HP0339		1	Y
			HP0428		1	Y
			HP0431	ptc1	1	Y
			HP0452		0.5	Y
			HP0677		1	Y
			HP0687	feoB	1	Y
			HP0775	spoT	1	Y
			HP0821	uvrC	1	Y
			HP0837		0.5	Y

		ID:12...72 GB:AE000511 percent identity: ...arity; putative	urease accessory protein
		ID:...47672 GB:AE000511 percent identity: ...rity; putative	urease accessory protein
		ID:...GB:AE000511 percent identity: ...ilarity; putative	urease accessory protein
		...1247...GB:AE000511 percent identity: ...arity; putative	urease accessory protein
		...01...E000511 percent identity: 100.00; ...ty; putative	urease beta subunit
		...00... PID:1567116 percent identity: 100.00; ...y; putative	urease alpha subunit- urea amidohydrolase
HP0075	ureC	similar to ...E000511 PID:2313156 percent identity: ...sequence similarity; putative	urease protein

Edges Patterns for Interaction



parent locus: uvrA
orf: HP0705
child locus:
orf: HP0677
probability: 1.0
exp. interaction: N

Solid red Line

Inferred data (Prob =1)

but not Exp data

parent locus: copA
orf: HP1072
child locus: mod
orf: HP0593
probability: 0.5
real interaction: N

Dotted Line

Inferred data only

parent locus: rpoB
orf: HP1198
child locus: trcF
orf: HP1541
probability: 1.0
exp. interaction: Y

Solid Line

Exp. Data and inferred Prob =1

parent locus: spoT
orf: HP0775
child locus: spoT
orf: HP0775
probability: 1.0
exp. interaction: Y

Solid curve -Self Interaction

Exp. Data and inferred Prob =1

parent locus:
orf: HP0677
child locus: uvrB
orf: HP1114
probability: 1.0
exp. interaction: N

Solid purple Line

(related with Level 2 object)

Inferred data (Prob =1)

but not Exp data

parent locus: uvrA
orf: HP0705
child locus:
orf: HP0452
probability: 0.5
exp. interaction: Y

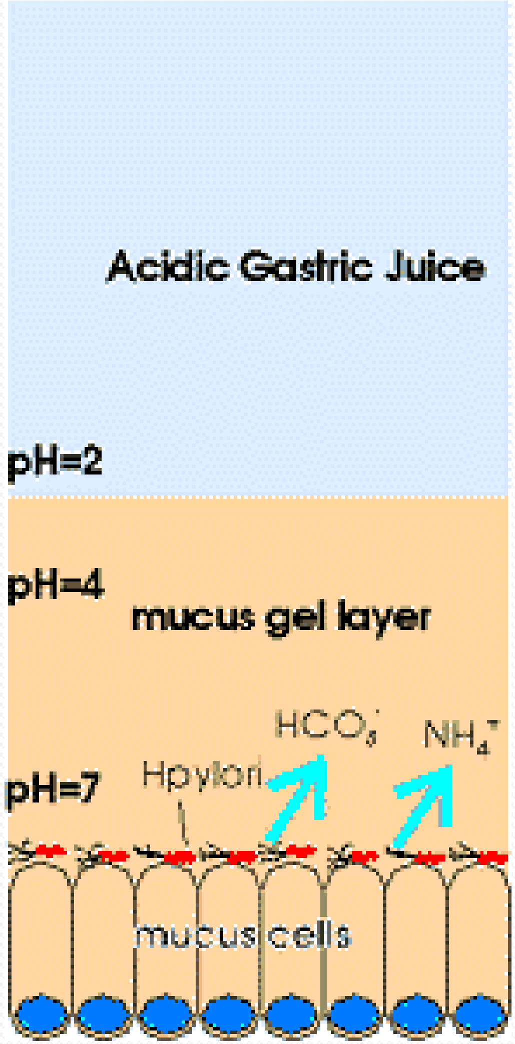
Dotted Line

Exp data

but Inferred data

With Prob<1

Discover New Research Targets with hp-DPI

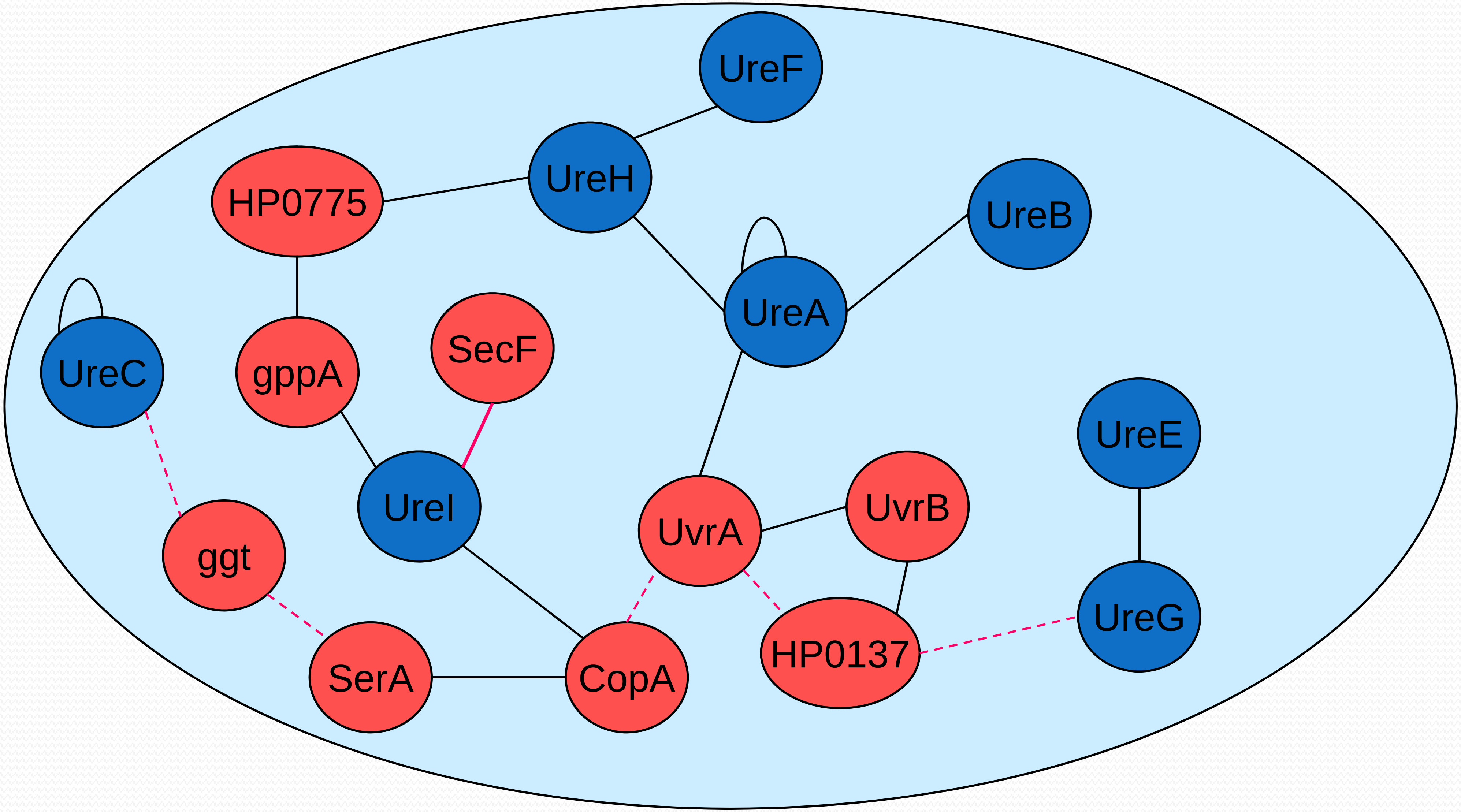


Urease

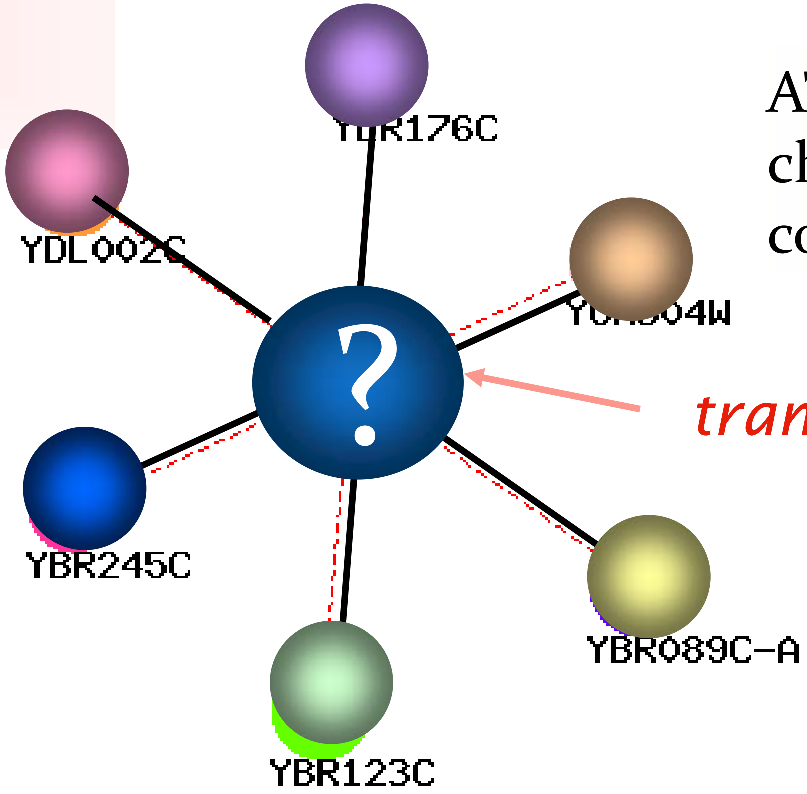
$$C=O(NH_2)2 + H^+ + 2H_2O \xrightarrow{\text{urease}} \boxed{HCO_3^- + 2\ NH_4^+}$$

bases

Network of Urease Complex



Annotated Protein Function by Interacting Partners



HMG1-box containing protein

DNA binding protein

ATPase component of chromatin remodeling complex

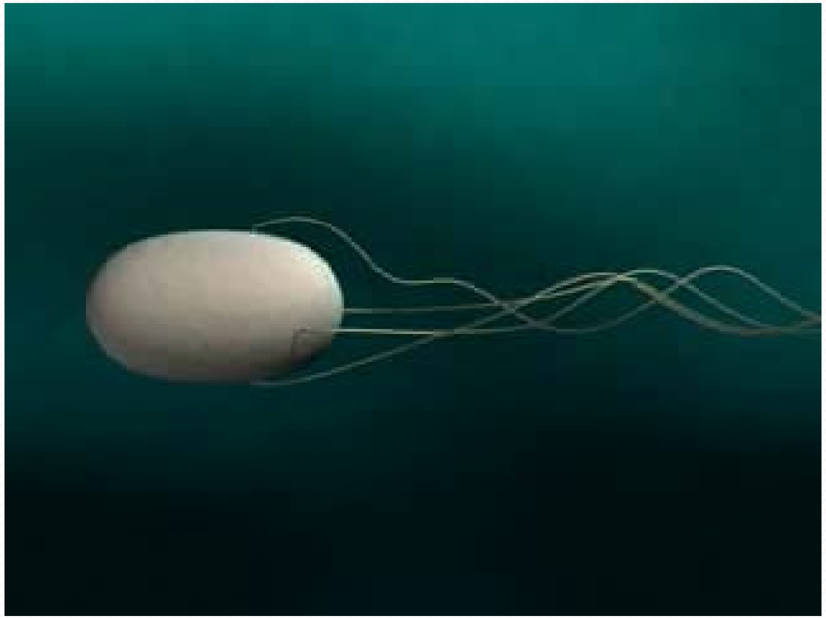
transcription related protein

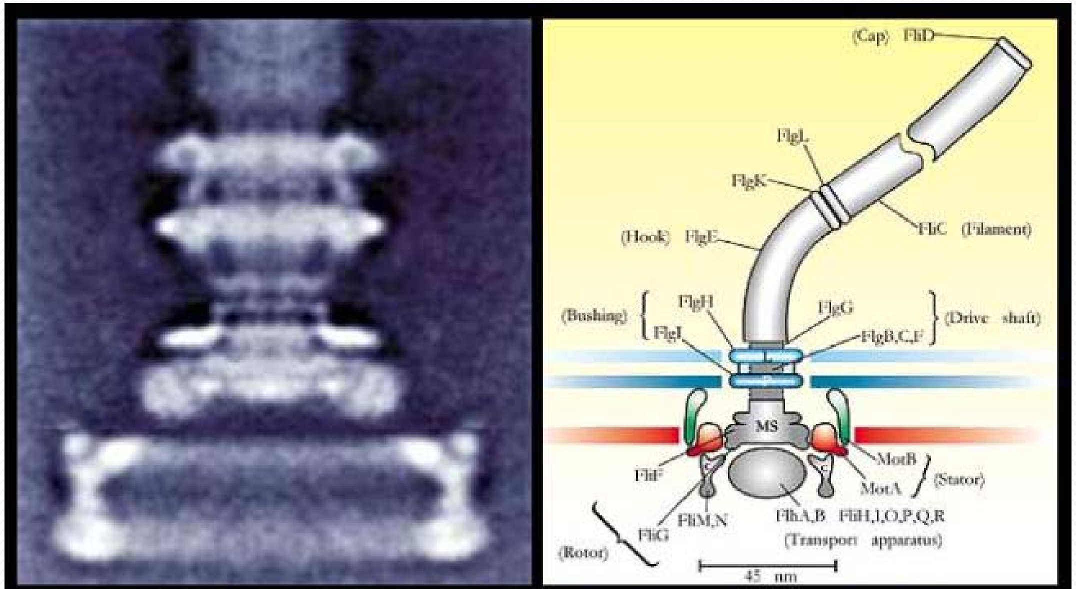
ATPase component of chromatin remodeling complex

Transcription regulation, chromatin binding

Transcription factor III, Tau

The Evolution of the Flagellum

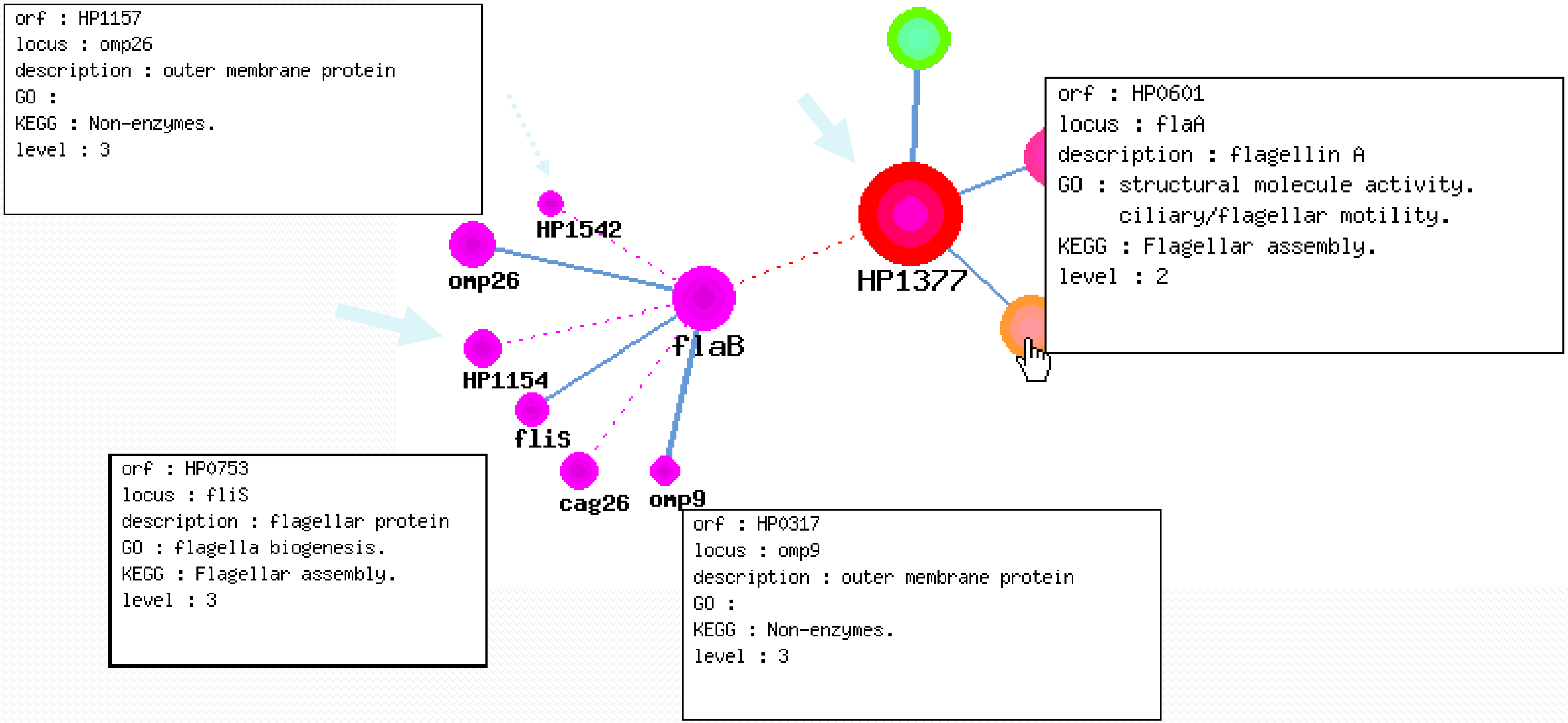




Structure of a bacterial flagellum. The illustration on the left is a rotationally averaged reconstruction of electron micrographs of purified hook-basal bodies. The names of the various parts are listed in the illustration to the right.

Reference: Uetz, et al., J. Bacteriol. 2006

Flagellum of H. pylori



orf : HP1157
locus : omp26
description : outer membrane protein
GO :
KEGG : Non-enzymes.
level : 3

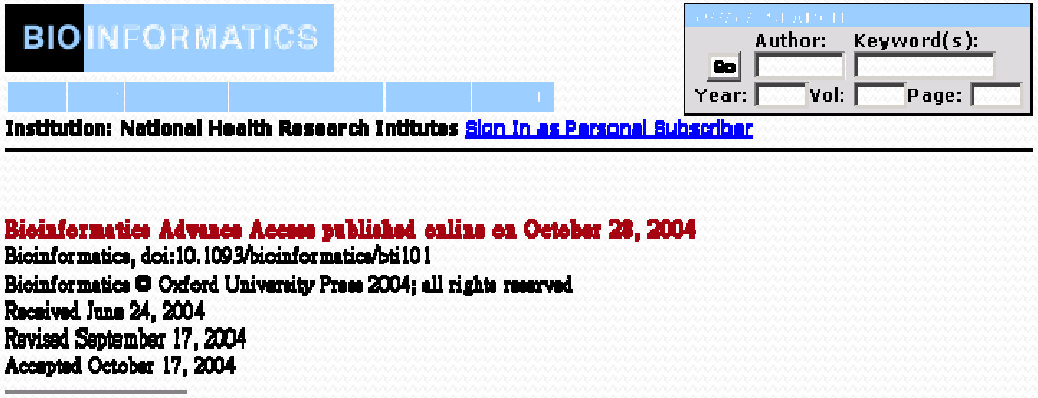
orf : HP0601
locus : flaA
description : flagellin A
GO : structural molecule activity.
ciliary/flagellar motility.
KEGG : Flagellar assembly.
level : 2

orf : HP0753
locus : fliS
description : flagellar protein
GO : flagella biogenesis.
KEGG : Flagellar assembly.
level : 3

orf : HP0317
locus : omp9
description : outer membrane protein
GO :
KEGG : Non-enzymes.
level : 3

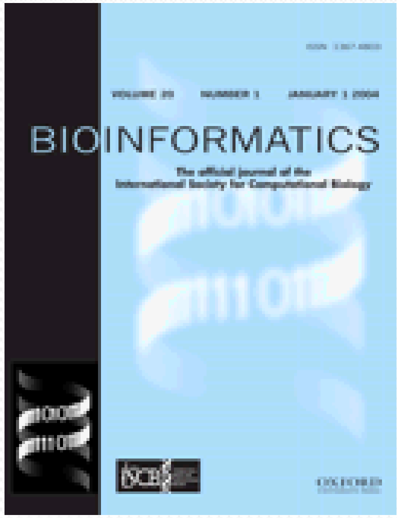
Reference: Uetz, et al., J. Bacteriol. 2006

hp-DPI (http://dpi.nhri.org.tw/hp/)



Institution: National Health Research Intitutes Sign In as Personal Subscriber

Bioinformatics Advance Access published online on October 28, 2004
Bioinformatics, doi:10.1093/bioinformatics/bti101
Bioinformatics © Oxford University Press 2004; all rights reserved
Received June 24, 2004
Revised September 17, 2004
Accepted October 17, 2004



BIOINFORMATICS
VOLUME 20 NUMBER 1 JANUARY 1 2004
The official journal of the International Society for Computational Biology
ISCB
OXFORD

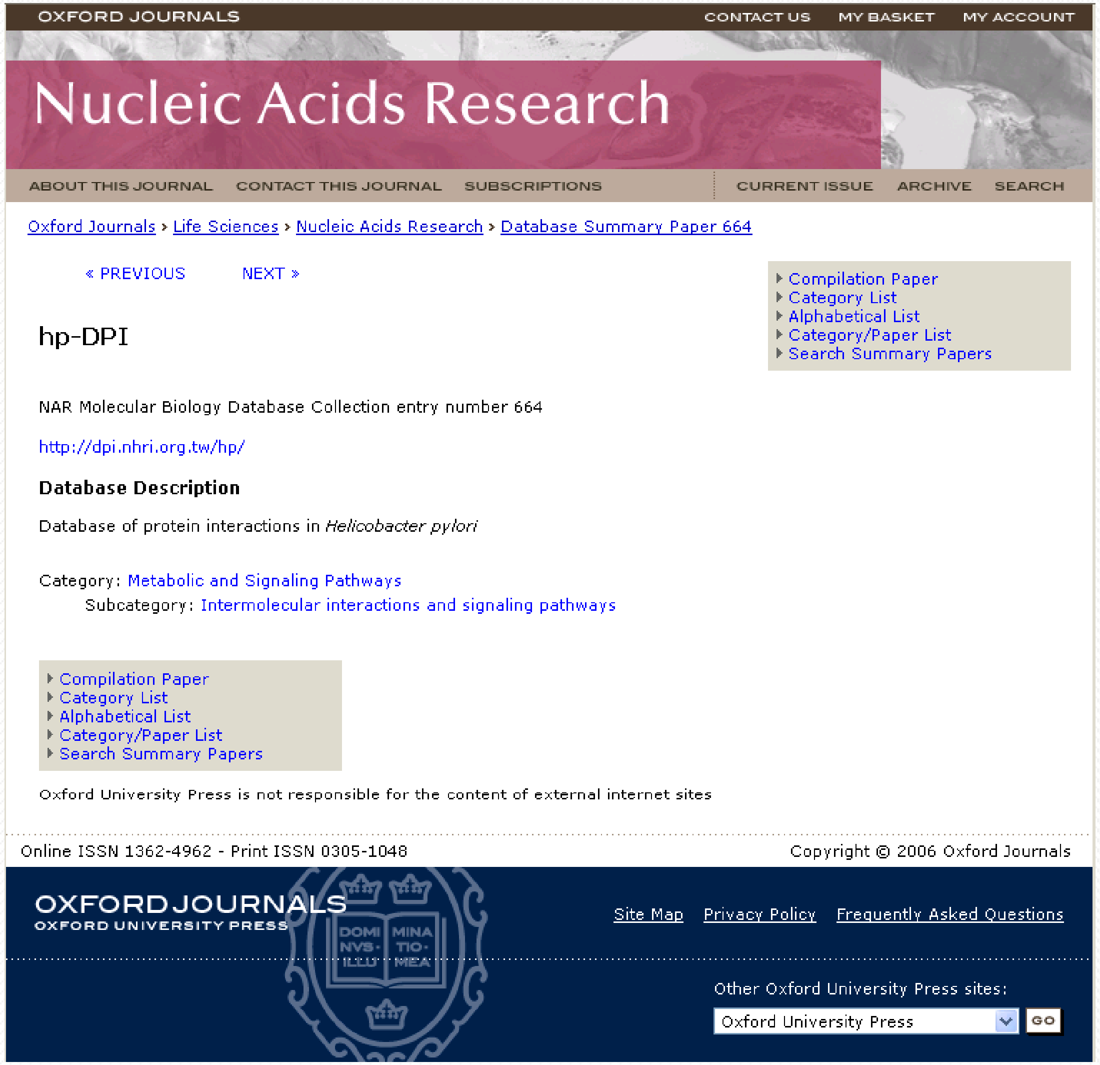
hp-DPI: *Helicobacter pylori* database of protein interactomes- embracing experimental and inferred interactions

Chung-Yen Lin [1]*, Chia-Ling Chen [1], Chi-Shiang Cho [1], Li-Ming Wang [1], Chia-Ming Chang [1], Pao-Yang Chen [1], Chen-Zen Lo [1], and Chao A. Hsiungs [1]

[1] Division of Biostatistics and Bioinformatics, National Health Research Institutes. #128, Sec. 2 Yaun-Chio-Yun Rd. Taipei 115, Taiwan

* To whom correspondence should be addressed.
Chung-Yen Lin, E-mail: cylin@nhri.org.tw

hp-DPI Selected into 2006 The Molecular Biology Database Collection by NAR



Oxford Journals > Life Sciences > Nucleic Acids Research > Database Summary Paper 664

hp-DPI

NAR Molecular Biology Database Collection entry number 664

http://dpi.nhri.org.tw/hp/

Database Description

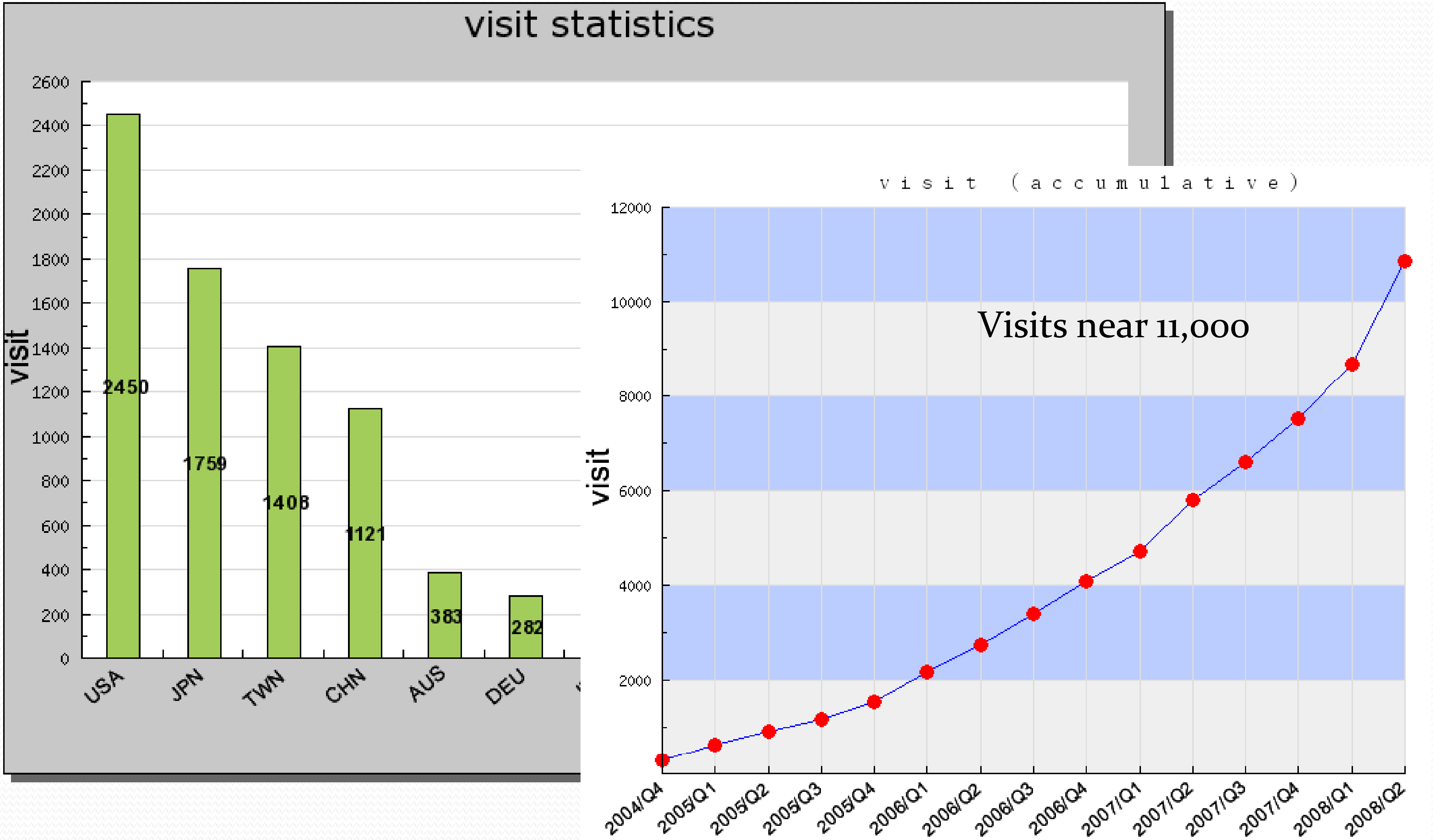
Database of protein interactions in *Helicobacter pylori*

Category: Metabolic and Signaling Pathways
Subcategory: Intermolecular interactions and signaling pathways

Oxford University Press is not responsible for the content of external internet sites

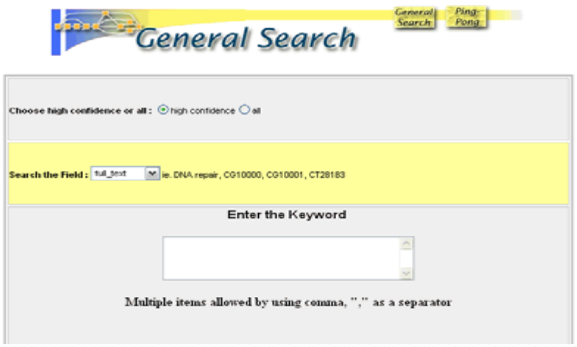
Online ISSN 1362-4962 - Print ISSN 0305-1048

Visit Statistics for hp-DPI from 2004/11/22 ~ 2008/07/1



Visits near 11,000

Previous Work II: Fly Database of Protein Interactomes



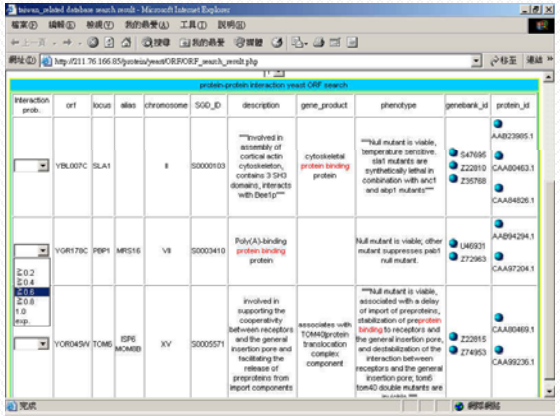
Choose high confidence or all : ◉ high confidence ○ all

Search the Field : full_text ie. DNA repair, CG10000, CG10001, CT28183

Enter the Keyword

Multiple items allowed by using comma, "," as a separator

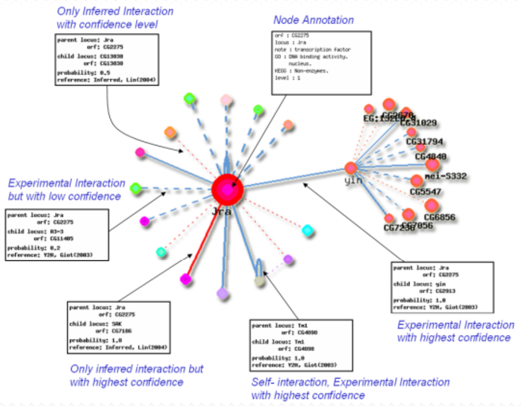
Query in Full Text


taiwan_related database search result - Microsoft Internet Explorer
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
上一頁 搜尋 我的最愛 媒體
網址(D) http://211.76.166.85/protein/yeast/ORF/ORF_search_result.php
移至 連結
≧0.2
≧0.4
≧0.6
≧0.8
1.0
exp.
完成
網際網路

protein-protein interaction yeast ORF search

interaction prob.	orf	locus	alias	chromosome	SGD_ID	description	gene_product	phenotype	genebank_id	protein_id
	YBL007C	SLA1		II	S0000103	"""Involved in assembly of cortical actin cytoskeleton, contains 3 SH3 domains, interacts with Bee1p"""	cytoskeletal protein binding protein	"""Null mutant is viable, temperature sensitive. sla1 mutants are synthetically lethal in combination with anc1 and abp1 mutants"""	S47695 Z22810 Z35768	AAB23985.1 CAA80463.1 CAA84826.1
	YGR178C	PBP1	MRS16	VII	S0003410	Poly(A)-binding protein binding protein		Null mutant is viable; other mutant suppresses pab1 null mutant.	U46931 Z72963	AAB94294.1 CAA97204.1
	YOR045W	TOM6	ISP6 MOM8B	XV	S0005571	involved in supporting the cooperativity between receptors and the general insertion pore and facilitating the release of preproteins from import components	associates with TOM40protein translocation complex component	"""Null mutant is viable, associated with a delay of import of preproteins, stabilization of preprotein binding to receptors and the general insertion pore, and destabilization of the interaction between receptors and the general insertion pore; tom6 tom40 double mutants are inviable."""	Z22815 Z74953	CAA80469.1 CAA99236.1

Search Result & Statistical Estimation


Only Inferred Interaction with confidence level
Node Annotation
Experimental Interaction but with low confidence
Only inferred interaction but with highest confidence
Self- interaction, Experimental Interaction with highest confidence
Experimental Interaction with highest confidence
Jra
yin
EG:152E8.4
CG2670
CG31029
CG31794
CG4840
mei-S332
CG5547
CG6856
CG7256
CG7056

parent locus: Jra
orf: CG2275
child locus: CG13838
orf: CG13838
probability: 0.5
reference: Inferred, Lin(2004)

orf : CG2275
locus : Jra
note : transcription factor
GO : DNA binding activity, nucleus.
KEGG : Non-enzymes.
level : 1

parent locus: Jra
orf: CG2275
child locus: R3-3
orf: CG11405
probability: 0.2
reference: Y2H, Giot(2003)

parent locus: Jra
orf: CG2275
child locus: SRK
orf: CG7186
probability: 1.0
reference: Inferred, Lin(2004)

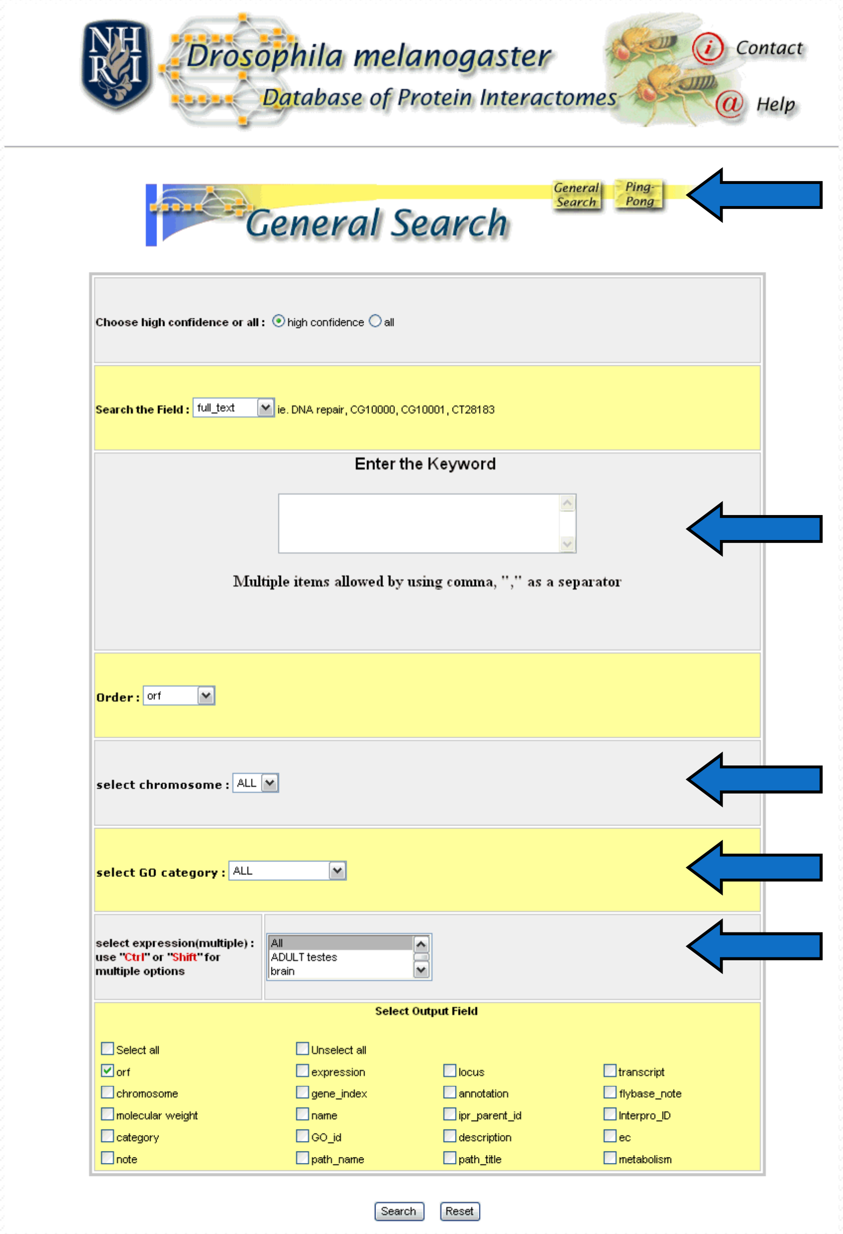
parent locus: Tm1
orf: CG4898
child locus: Tm1
orf: CG4898
probability: 1.0
reference: Y2H, Giot(2003)

parent locus: Jra
orf: CG2275
child locus: yin
orf: CG2913
probability: 1.0
reference: Y2H, Giot(2003)

Network Visualization & popup annotation

http://flydpi.nhri.og.tw

New Features in FlyDPI



General Search

Choose high confidence or all : ● high confidence ○ all

Search the Field : full_text ie. DNA repair, CG10000, CG10001, CT28183

Enter the Keyword

Multiple items allowed by using comma, "," as a separator

Order : orf

select chromosome : ALL

select GO category : ALL

select expression(multiple) : use "Ctrl" or "Shift" for multiple options

All
ADULT testes
brain

Select Output Field

□ Select all	□ Unselect all		
☑ orf	□ expression	□ locus	□ transcript
□ chromosome	□ gene_index	□ annotation	□ flybase_note
□ molecular weight	□ name	□ ipr_parent_id	□ Interpro_ID
□ category	□ GO_id	□ description	□ ec
□ note	□ path_name	□ path_title	□ metabolism

Ping-Pong Search

Full-text Search

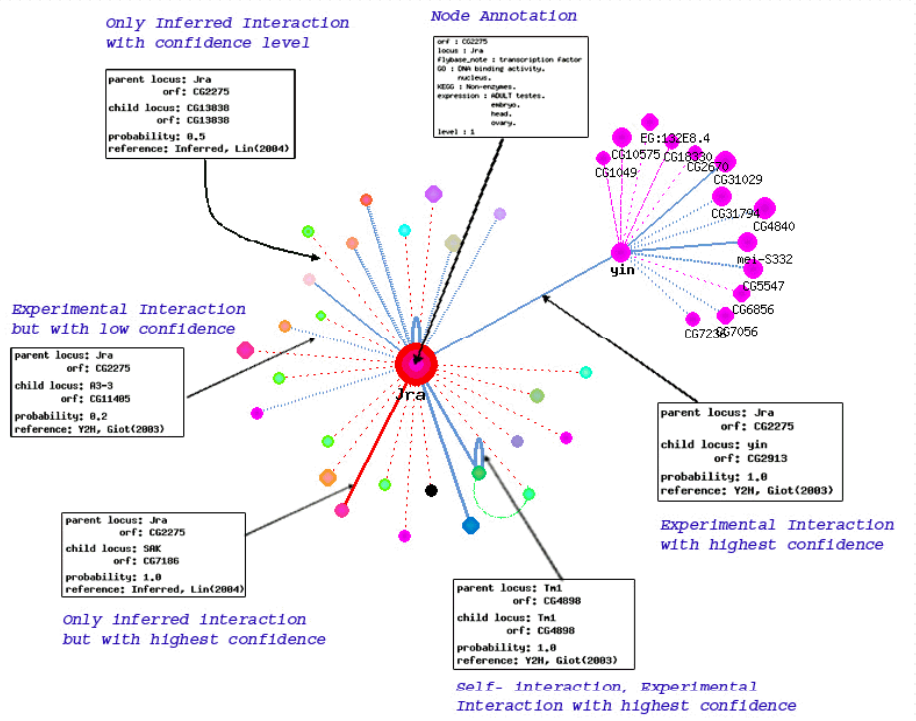
Chromosome Location

Gene Categories form GO

Spatiotemporal Scenarios

Search Results of FlyDPI

General Search



Only Inferred Interaction with confidence level

```
parent locus: Jra
         orf: CG2275
child locus: CG13838
        orf: CG13838
probability: 0.5
reference: Inferred, Lin(2004)
```

Node Annotation

```
orf : CG2275
locus : Jra
Flybase_note : transcription factor
GO : DNA binding activity.
     nucleus.
KEGG : Non-enzymes.
expression : ADULT testes.
             embryo.
             head.
             ovary.
level : 1
```

Experimental Interaction but with low confidence

```
parent locus: Jra
         orf: CG2275
child locus: A3-3
        orf: CG11405
probability: 0.2
reference: Y2H, Giot(2003)
```

```
parent locus: Jra
         orf: CG2275
child locus: yin
        orf: CG2913
probability: 1.0
reference: Y2H, Giot(2003)
```

Experimental Interaction with highest confidence

```
parent locus: Jra
         orf: CG2275
child locus: SAK
        orf: CG7186
probability: 1.0
reference: Inferred, Lin(2004)
```

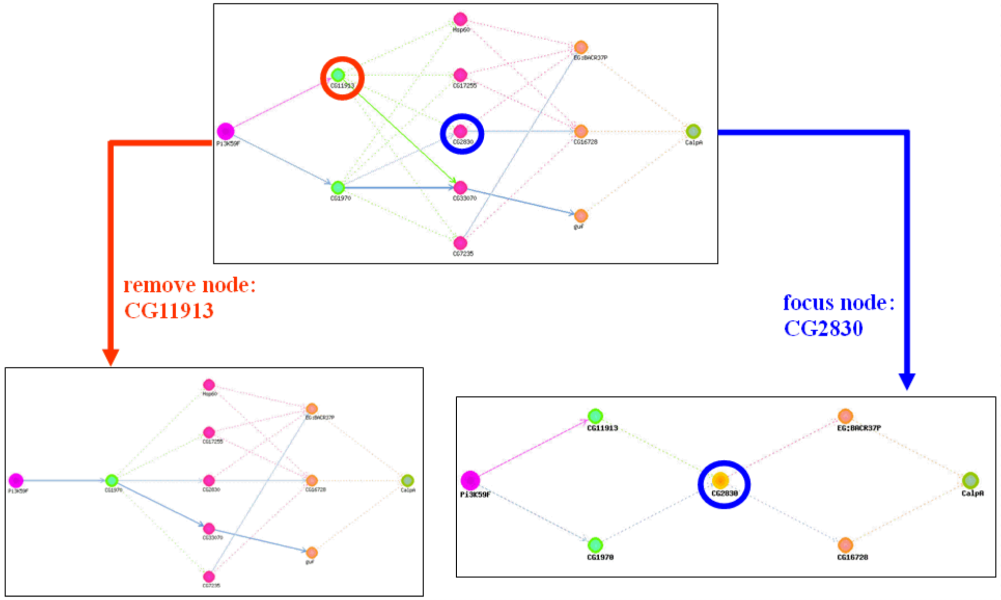
Only inferred interaction but with highest confidence

```
parent locus: Tm1
         orf: CG4898
child locus: Tm1
        orf: CG4898
probability: 1.0
reference: Y2H, Giot(2003)
```

Self- interaction, Experimental Interaction with highest confidence

A snap of the experimental and inferred visualized interaction networks of *D. melanogaster* interactome under specific spatiotemporal scenarios.

Ping-Pong Search

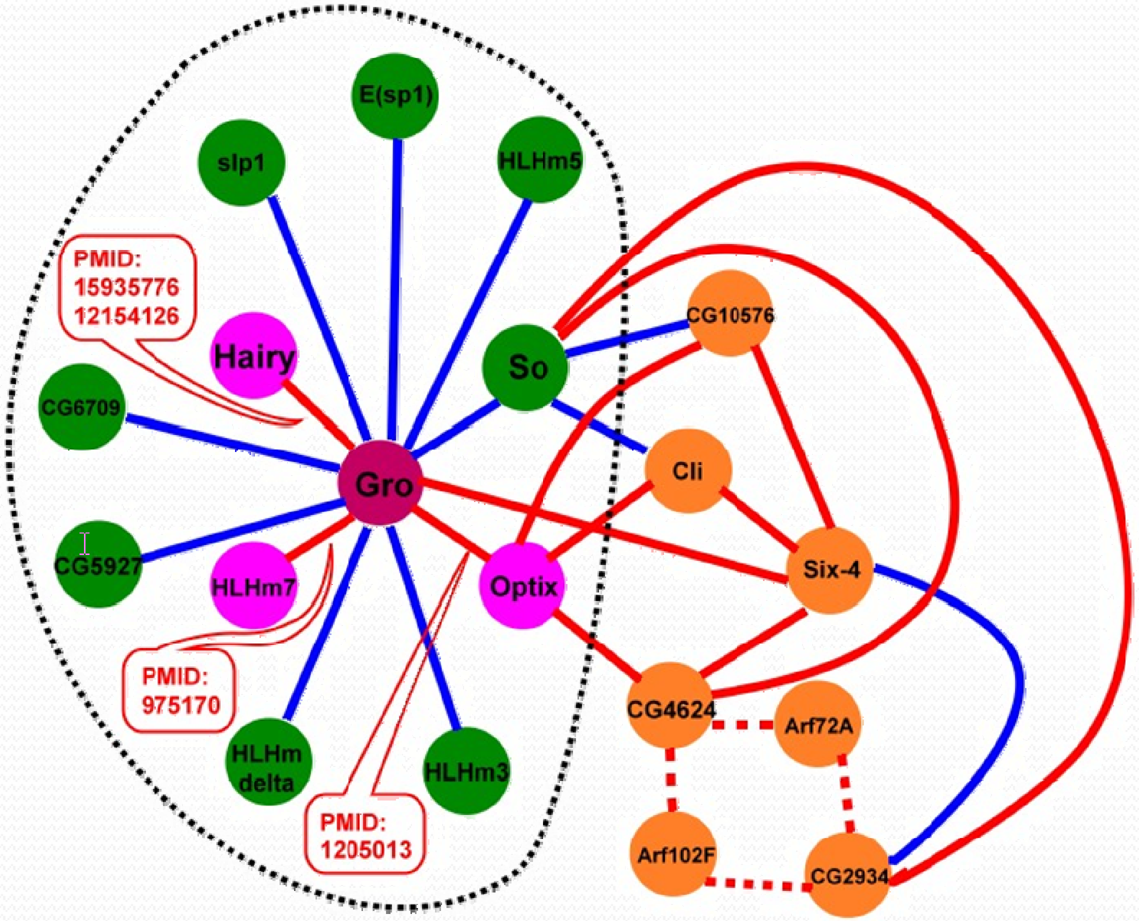


remove node:
CG11913

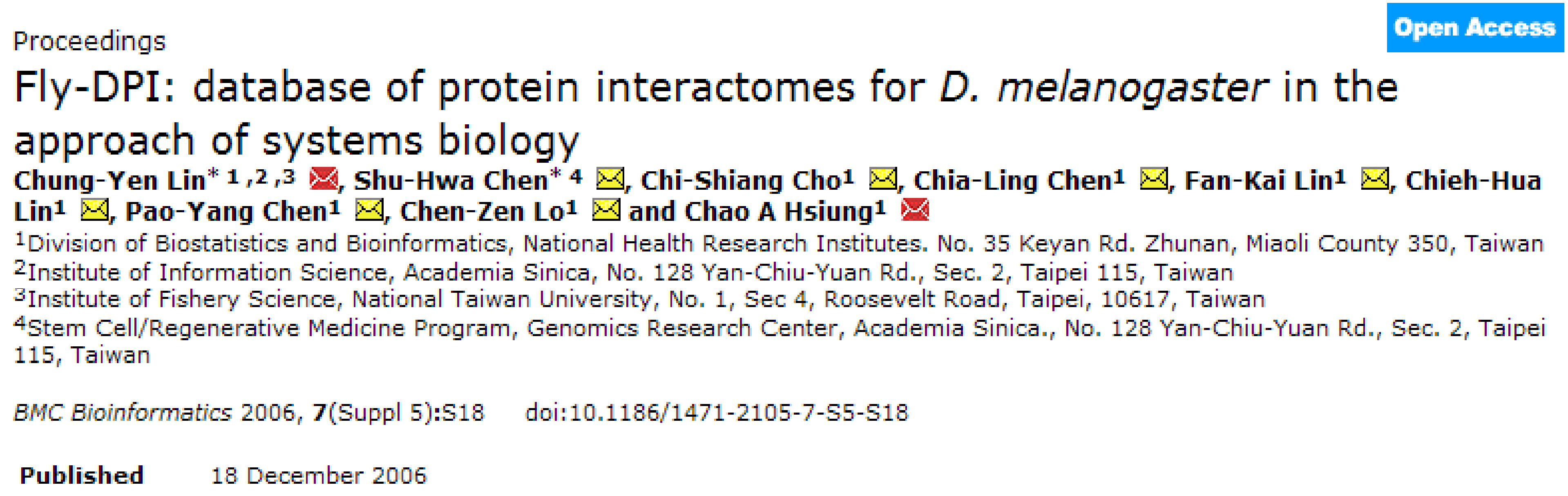
focus node:
CG2830

Map of proteins potentially involved in apoptosis generated by ping-pong search. By the click on the nodes or lines between two query proteins, the advanced option will remove the paths related or confine the paths with the selected nodes or lines

Interaction Network Amid Gro And Its Partners



FlyDPI -- (http://flydpi.nhri.org.tw)



Proceedings

Open Access

Fly-DPI: database of protein interactomes for *D. melanogaster* in the approach of systems biology

Chung-Yen Lin* [1,2,3] ✉, **Shu-Hwa Chen*** [4] ✉, **Chi-Shiang Cho**[1] ✉, **Chia-Ling Chen**[1] ✉, **Fan-Kai Lin**[1] ✉, **Chieh-Hua Lin**[1] ✉, **Pao-Yang Chen**[1] ✉, **Chen-Zen Lo**[1] ✉ **and Chao A Hsiung**[1] ✉

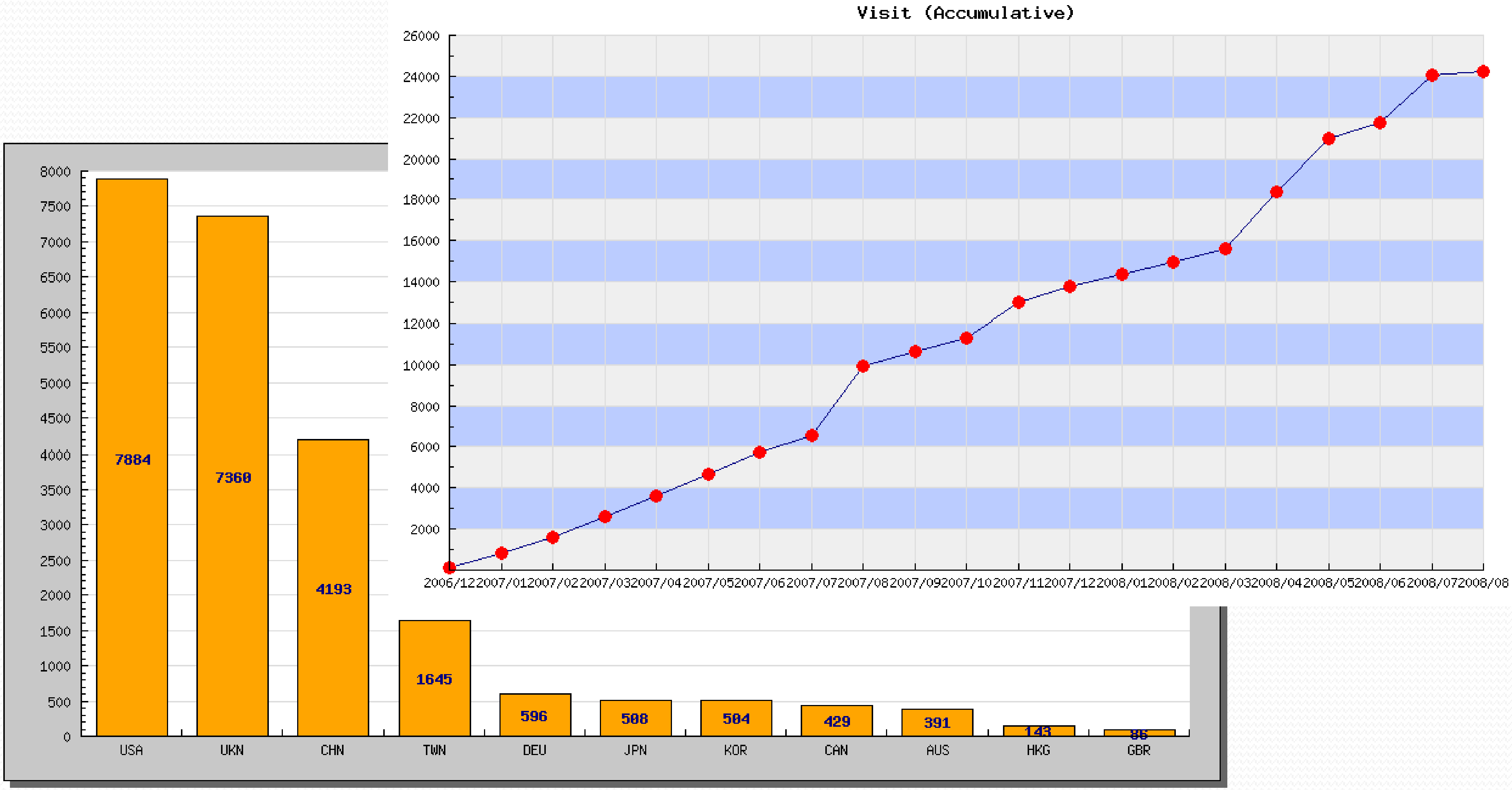
[1]Division of Biostatistics and Bioinformatics, National Health Research Institutes. No. 35 Keyan Rd. Zhunan, Miaoli County 350, Taiwan
[2]Institute of Information Science, Academia Sinica, No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan
[3]Institute of Fishery Science, National Taiwan University, No. 1, Sec 4, Roosevelt Road, Taipei, 10617, Taiwan
[4]Stem Cell/Regenerative Medicine Program, Genomics Research Center, Academia Sinica., No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan

BMC Bioinformatics 2006, **7**(Suppl 5):S18   doi:10.1186/1471-2105-7-S5-S18

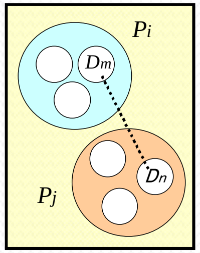
Published 18 December 2006

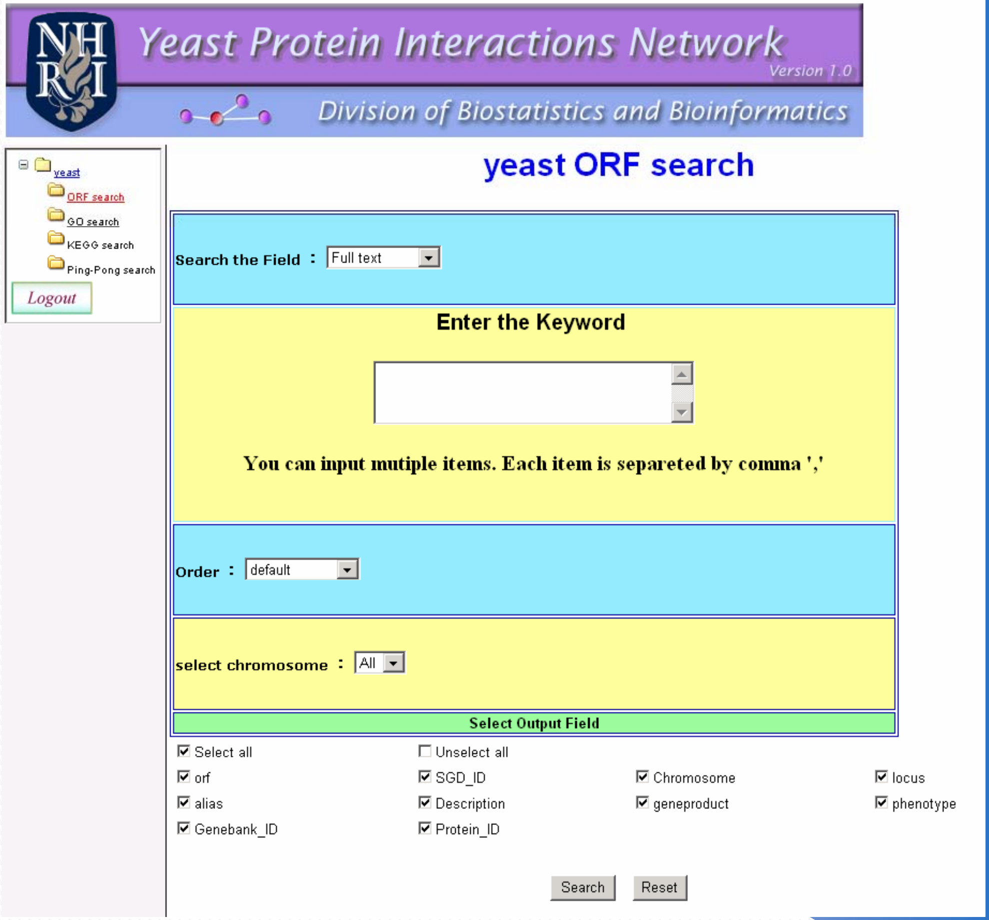
BMC
Bioinformatics

Visits of FlyDPI *(Dec 2006- Aug 2008)*



Framework for Database of Protein Interactome, DPI




NHRI
Yeast Protein Interactions Network
Version 1.0
Division of Biostatistics and Bioinformatics
yeast
ORF search
GO search
KEGG search
Ping-Pong search
Logout

yeast ORF search

Search the Field : Full text

Enter the Keyword

You can input mutiple items. Each item is separeted by comma ','

Order : default

select chromosome : All

Select Output Field

☑ Select all ☐ Unselect all

☑ orf ☑ SGD_ID ☑ Chromosome ☑ locus

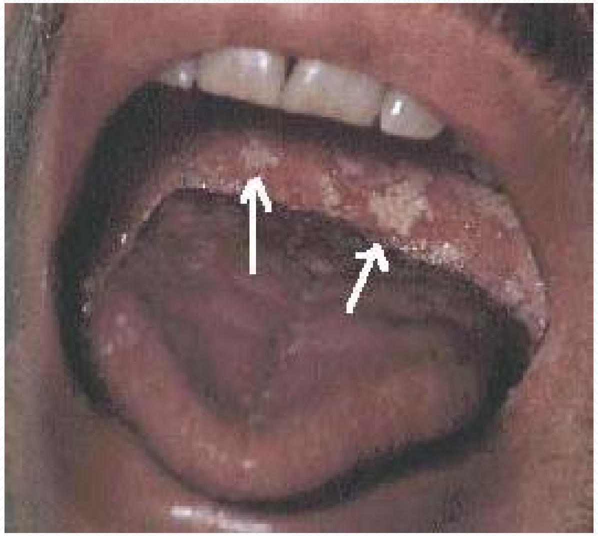
☑ alias ☑ Description ☑ geneproduct ☑ phenotype

☑ Genebank_ID ☑ Protein_ID

Search Reset

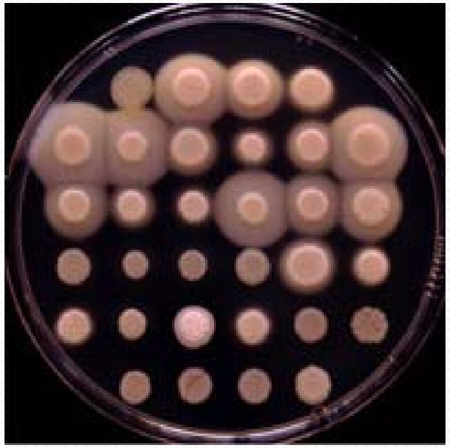
Candida albicans Protein Network

- *Candida albicans* is both a commensal and pathogen of humans that can infect a broad range of body sites.
- Endogenous *C. albicans* infections are established by cells that normally colonise mucosal surfaces or skin as harmless commensals, and that are triggered to cause infection by changes in the host immune system or microflora.

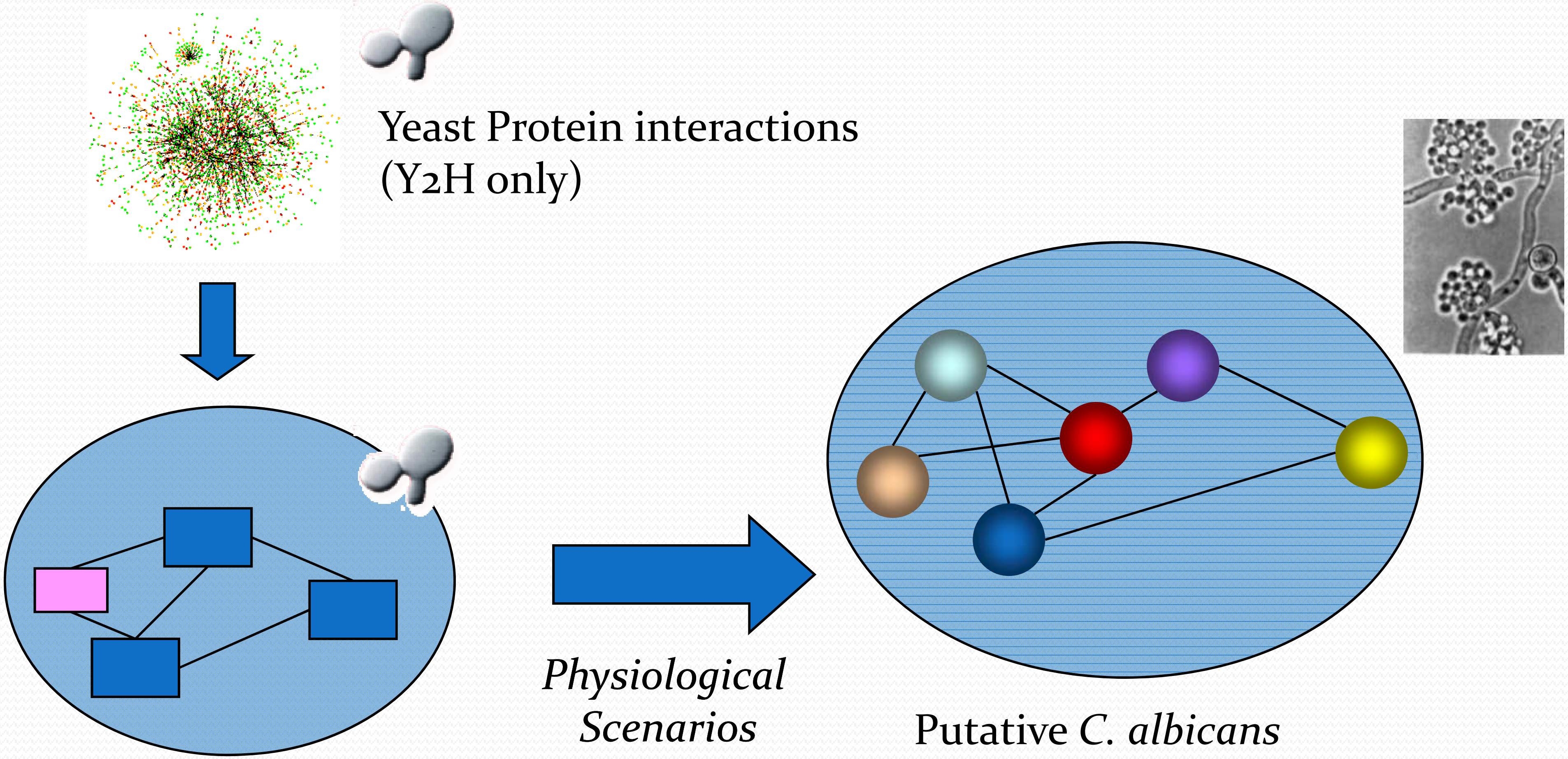


C. Albicans with Antifungal Drug Resistance

- Question emerging after abusing Antifungal drugs
- Identification of novel drug targets by network biology is way to solve the problem.



Inferred Protein Interactions by hidden DDIs from Yeast



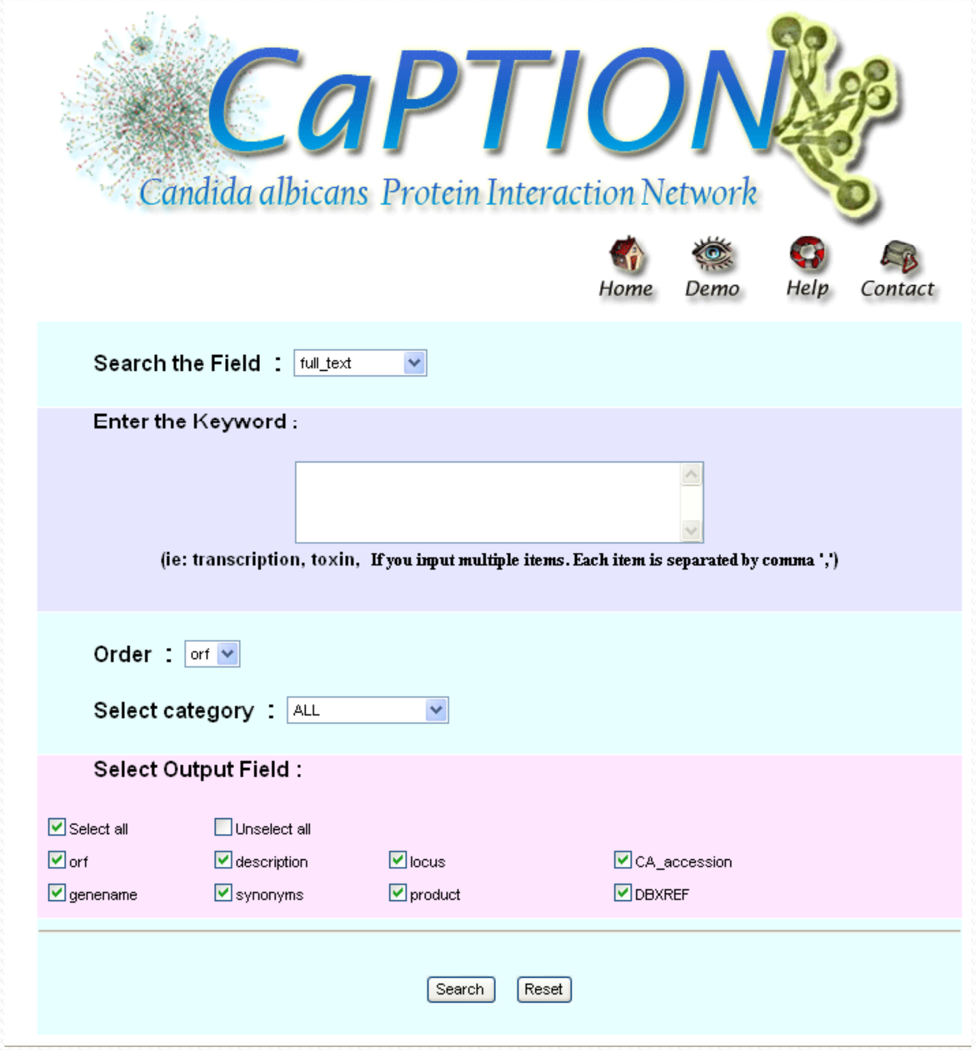
Yeast Protein interactions (Y2H only)

Physiological Scenarios

Putative *C. albicans* Protein interaction

Domain-interaction Network of yeast

CaPTION-
C. albicans Protein interaction Network



Search the Field : full_text

Enter the Keyword :

(ie: transcription, toxin, If you input multiple items. Each item is separated by comma ',')

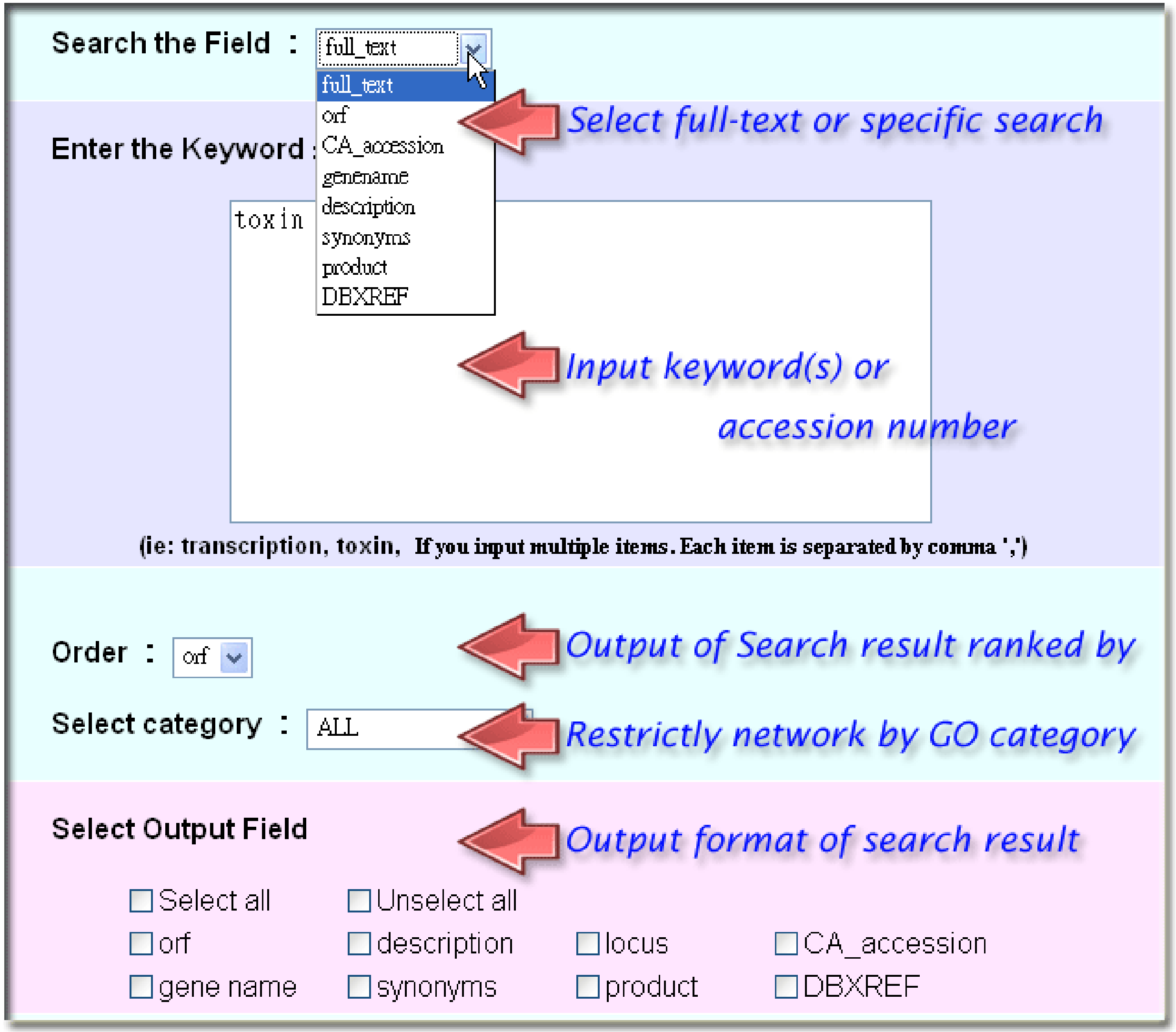
Order : orf

Select category : ALL

Select Output Field :

☑ Select all ☐ Unselect all
☑ orf ☑ description ☑ locus ☑ CA_accession
☑ genename ☑ synonyms ☑ product ☑ DBXREF

Interface of CaPTION (available soon)



Search the Field ： full_text

Select full-text or specific search

Enter the Keyword : toxin

Input keyword(s) or accession number

(ie: transcription, toxin, If you input multiple items. Each item is separated by comma ',')

Order ： orf

Output of Search result ranked by

Select category ： ALL

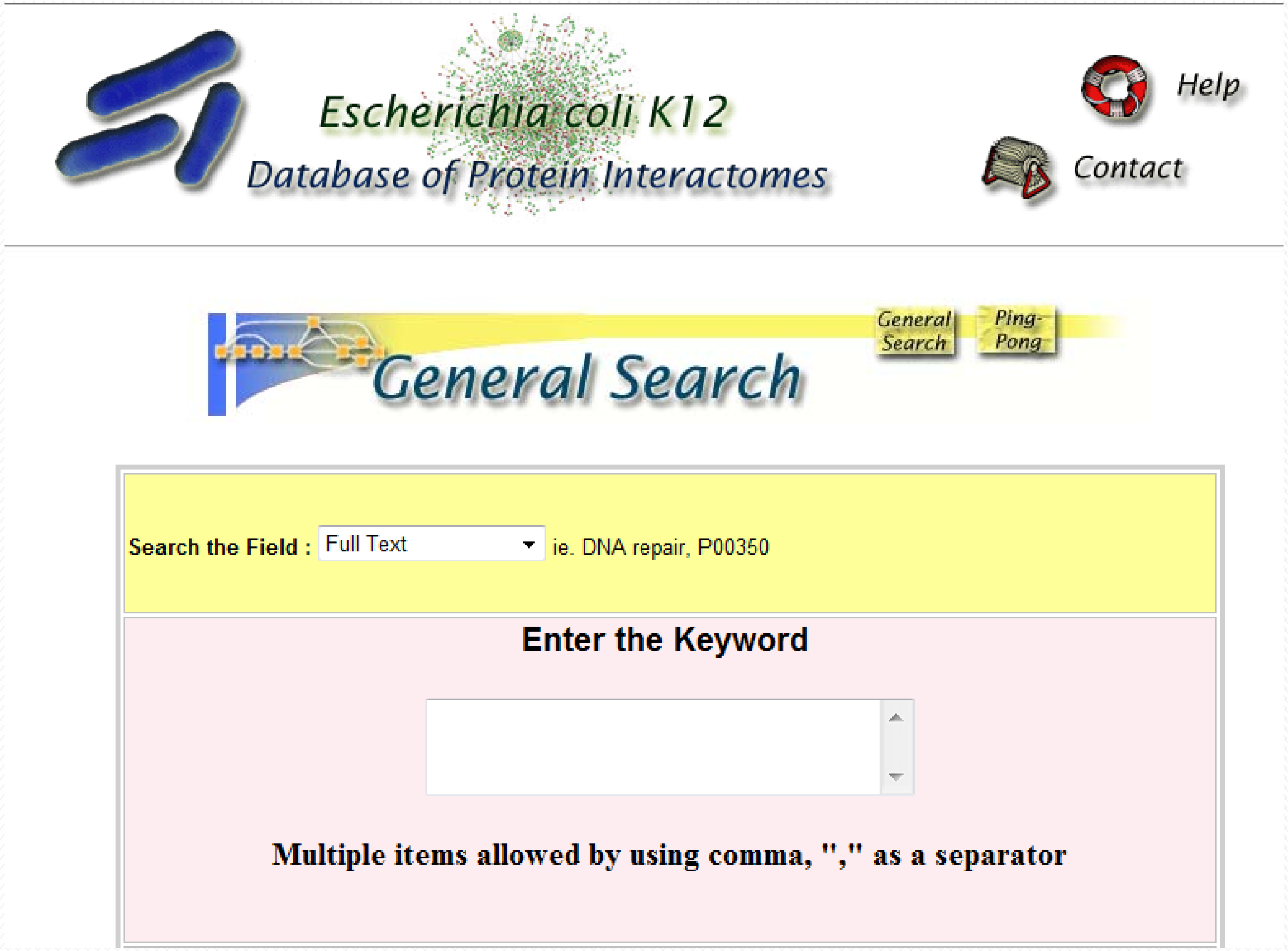
Restrictly network by GO category

Select Output Field

Output format of search result

□ Select all	□ Unselect all		
□ orf	□ description	□ locus	□ CA_accession
□ gene name	□ synonyms	□ product	□ DBXREF

Ecoli-DPI

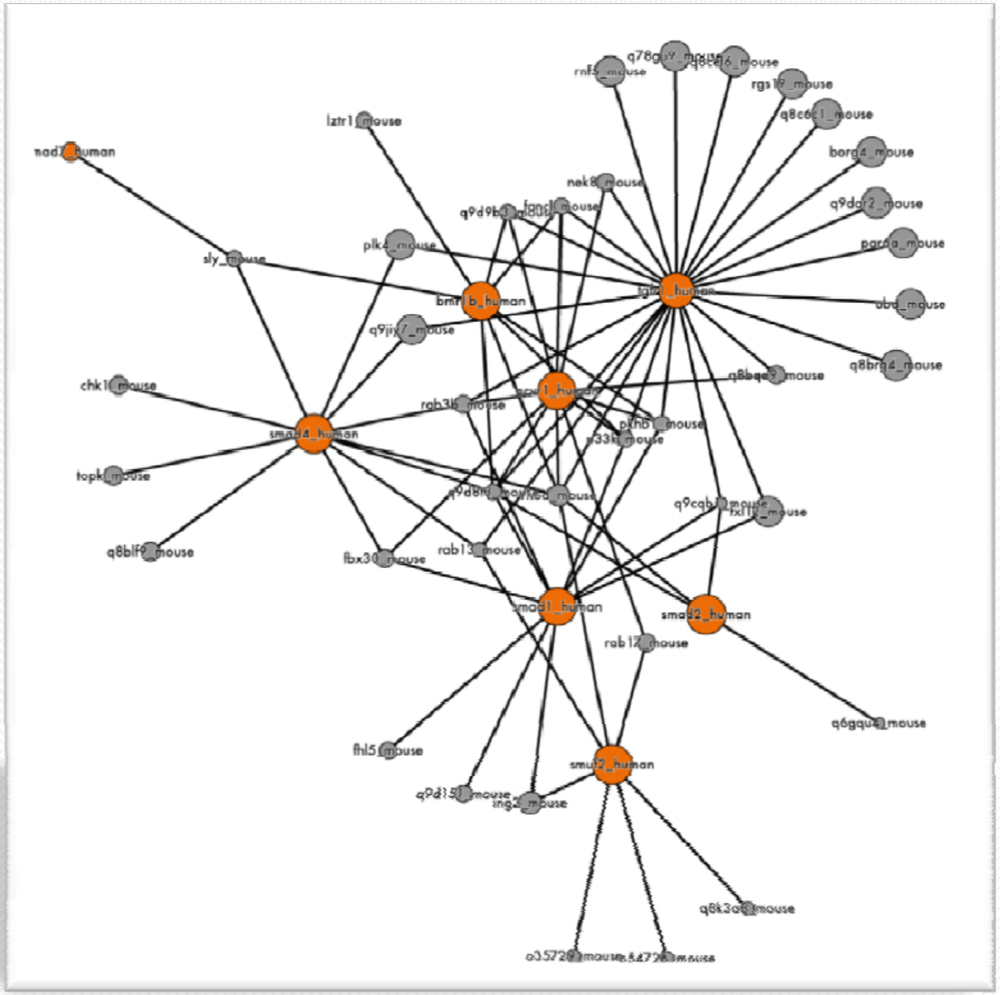


Search the Field : Full Text ie. DNA repair, P00350

Enter the Keyword

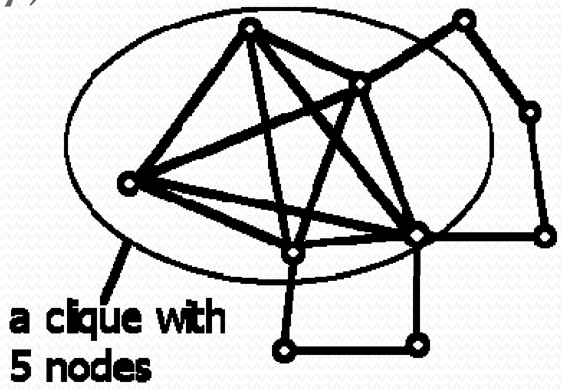
Multiple items allowed by using comma, "," as a separator

Network Biology: Hub/ Essential Proteins Identification

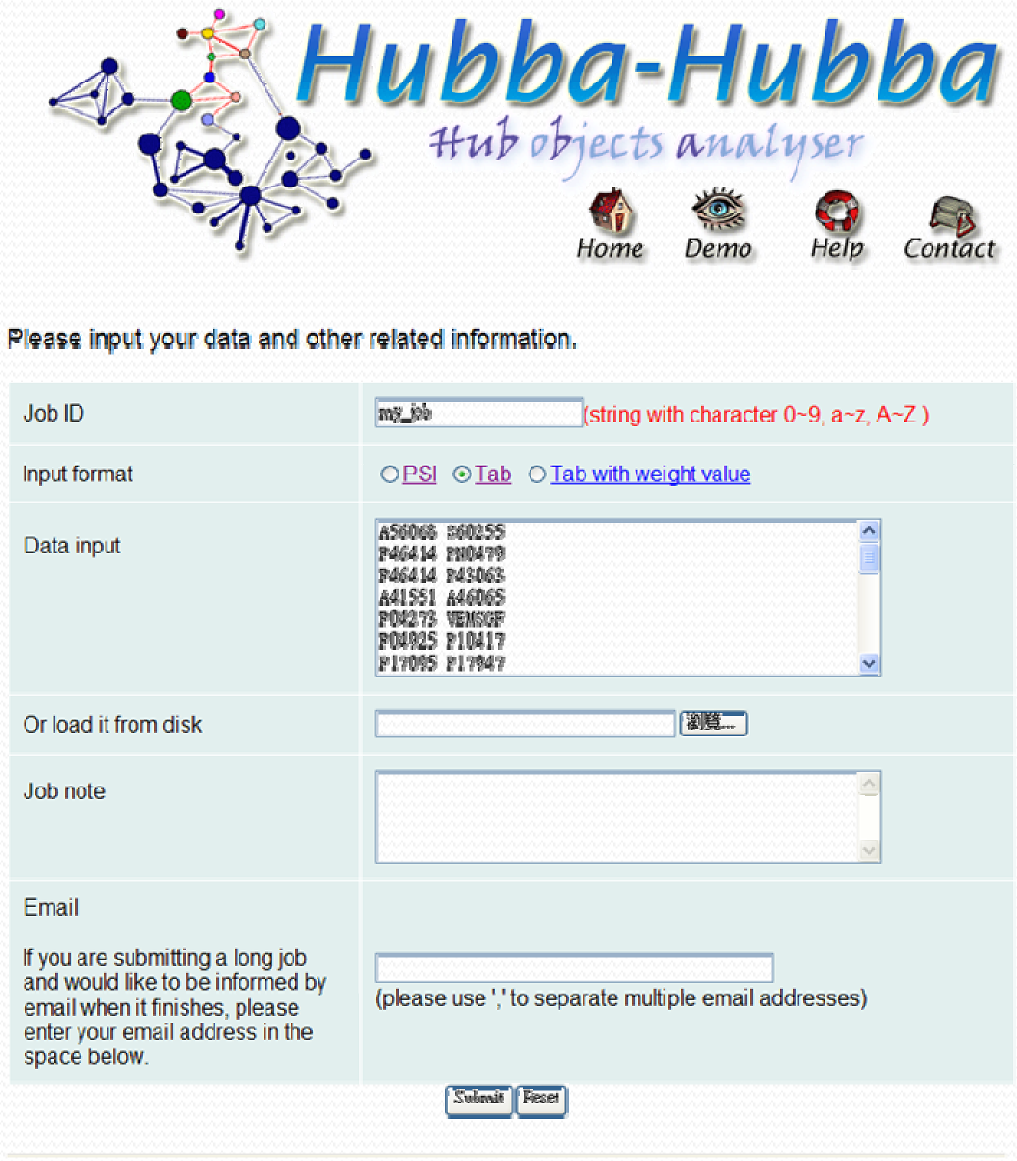


The Ways to Detect Hubs

- Degree (Jeong H. *et al.*,2001)
- Bottle Neck (Przulj N. *et al.*, 2003)
- Percolation Based **(Vi)** (Chin *et al.*, 2003)
- Subgraph centrality **(SC)** (Ernesto, E *et al.*, 2005)
- Maximum Connected Component from Neighborhood Induced Subgraph (**MNCIS**) (Our team, 2007)
- Maximum Connected Component from Neighborhood Induced Subgraph with Density **(MNCISD)** (Our team, 2007)



Hub Object Analyzer: Hubba

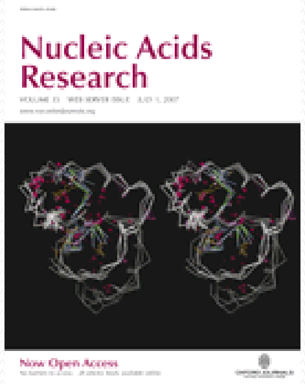


Please input your data and other related information.

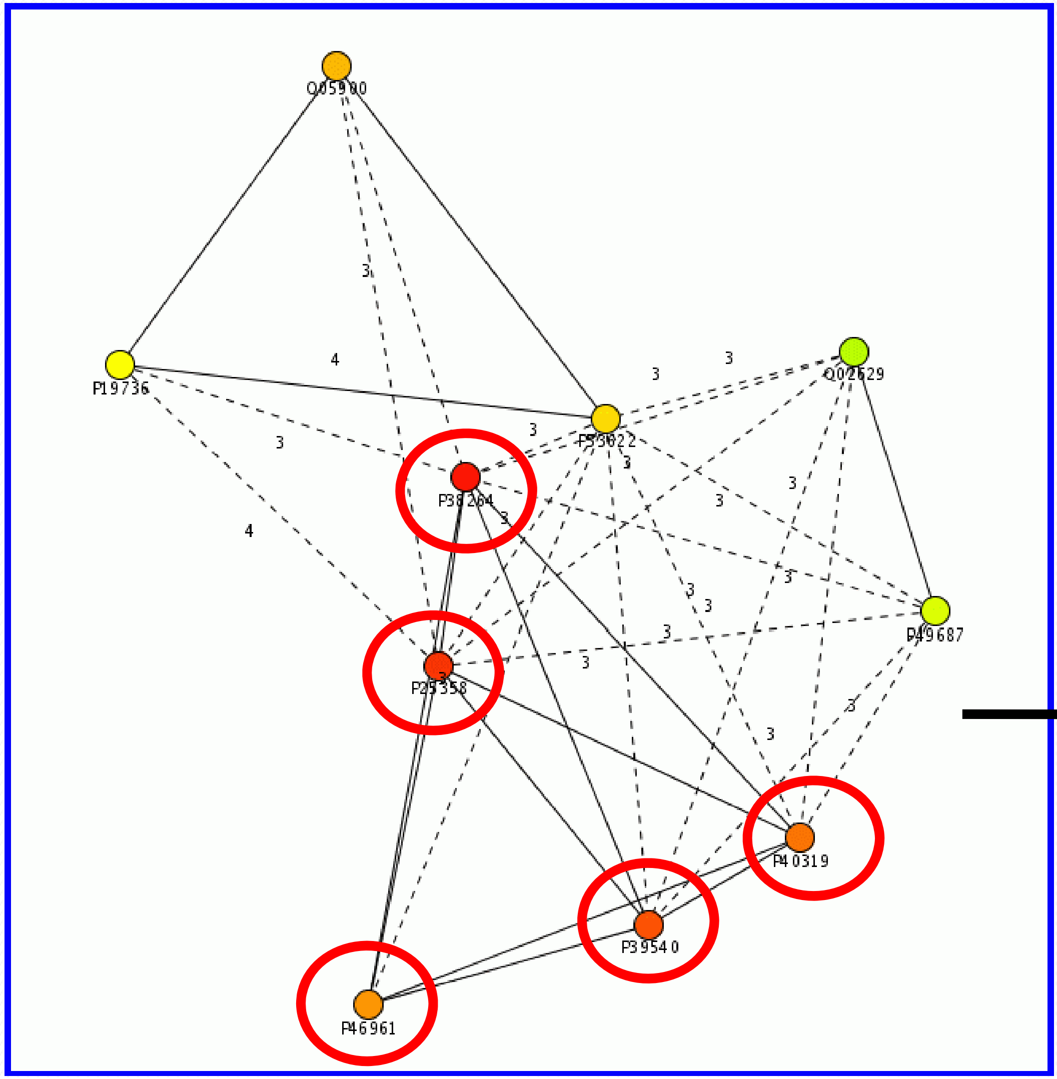
Job ID	my_job (string with character 0~9, a~z, A~Z)
Input format	○ PSI ● Tab ○ Tab with weight value
Data input	A56068 S60255 P46414 PN0479 P46414 P43063 A41551 A46065 P04273 VEMSGF P04925 P10417 P17095 P17947
Or load it from disk	瀏覽...
Job note	
Email If you are submitting a long job and would like to be informed by email when it finishes, please enter your email address in the space below.	(please use ',' to separate multiple email addresses)

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

To appear on NAR 2008 Web issue

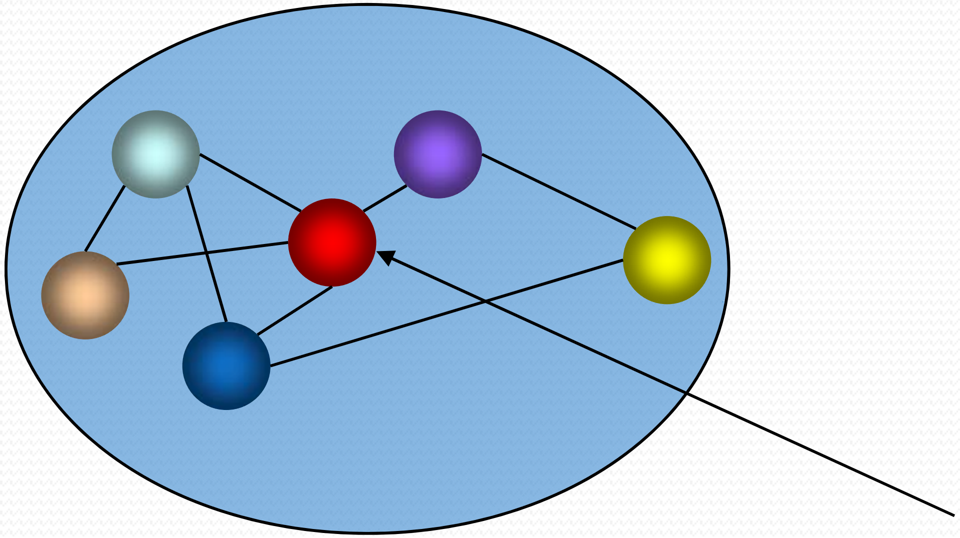


The Relationship of Top 10 in Yeast Complex Network (PPI from DIP, 2007 Jan)



Fragile motif in whole network

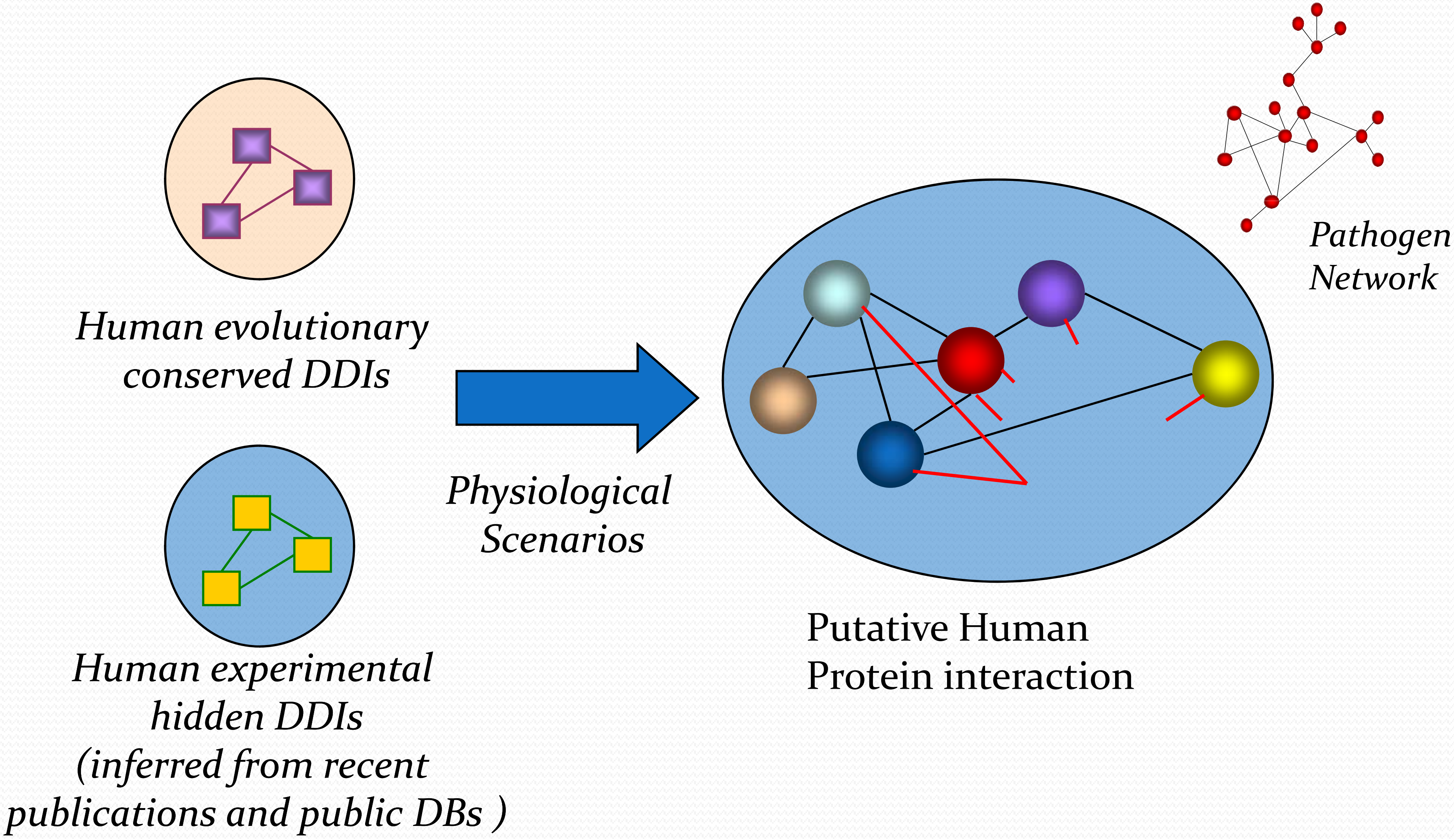
Identification Target Proteins and Hubs for Novel Cancer Therapies



Putative Protein Network in Human Cancer

Hub protein can be treated with RNA inference to perturb the network, then stop the progress of tumors

Inferred Protein Interactions by Conserved and hidden DDIs



Human evolutionary conserved DDIs

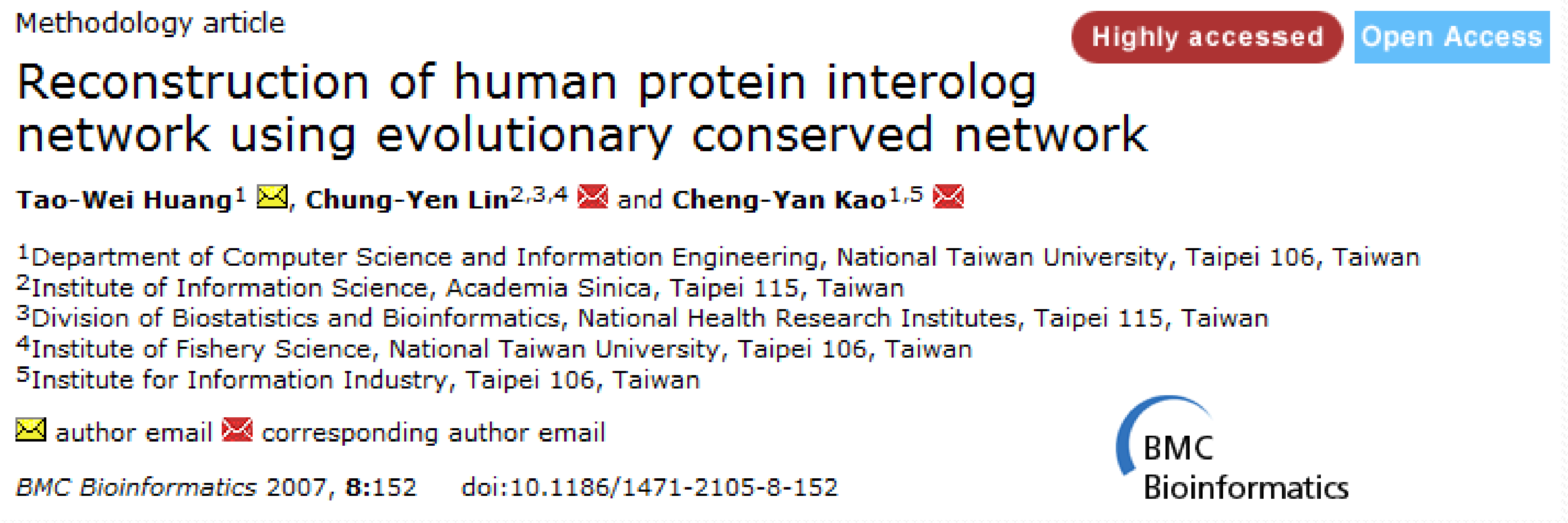
Human experimental hidden DDIs (inferred from recent publications and public DBs)

Physiological Scenarios

Putative Human Protein interaction

Pathogen Network

Human Protein Network



Methodology article

Highly accessed Open Access

Reconstruction of human protein interolog network using evolutionary conserved network

Tao-Wei Huang[1] ✉, **Chung-Yen Lin**[2,3,4] ✉ and **Cheng-Yan Kao**[1,5] ✉

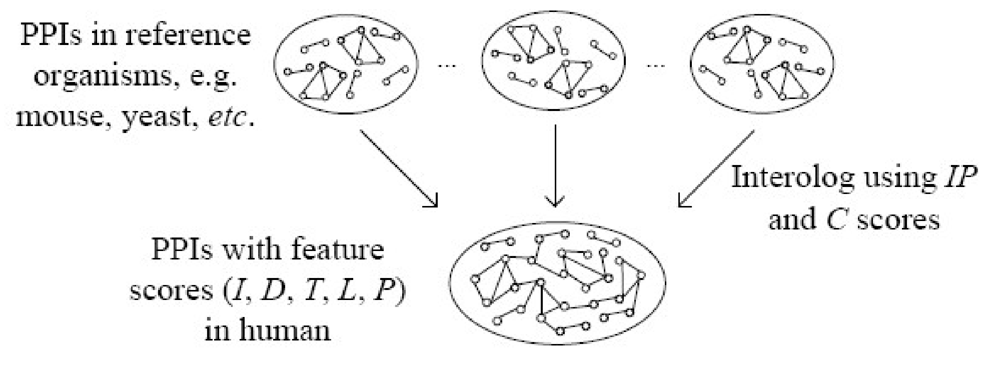
[1] Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan
[2] Institute of Information Science, Academia Sinica, Taipei 115, Taiwan
[3] Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taipei 115, Taiwan
[4] Institute of Fishery Science, National Taiwan University, Taipei 106, Taiwan
[5] Institute for Information Industry, Taipei 106, Taiwan

✉ author email ✉ corresponding author email

BMC Bioinformatics 2007, **8**:152 doi:10.1186/1471-2105-8-152

BMC Bioinformatics

$$CS = w_I * \frac{I}{I_R} + w_D * \frac{D}{D_K} + w_T * \frac{T}{T_K} + w_L * \frac{L}{L_K} + w_P * \frac{P}{P_K}$$



PPIs in reference organisms, e.g. mouse, yeast, *etc*.
Interolog using *IP* and *C* scores
PPIs with feature scores (*I*, *D*, *T*, *L*, *P*) in human

Confidence score (CS)

- Interolog score (I)

$$I_{ij} = w_{ec} * \min(IP_{A_i}, IP_{B_j}) * C_{ab}$$

- Domain-domain combination score (D)

$$D = \sum_{j=1}^{2^m-1} \sum_{i=1}^{2^m-1} \frac{N'(pd_i, pd_j)}{N(pd_i, pd_j)} \text{ if } pd_i \in PD_d, pd_j \in PD_d$$

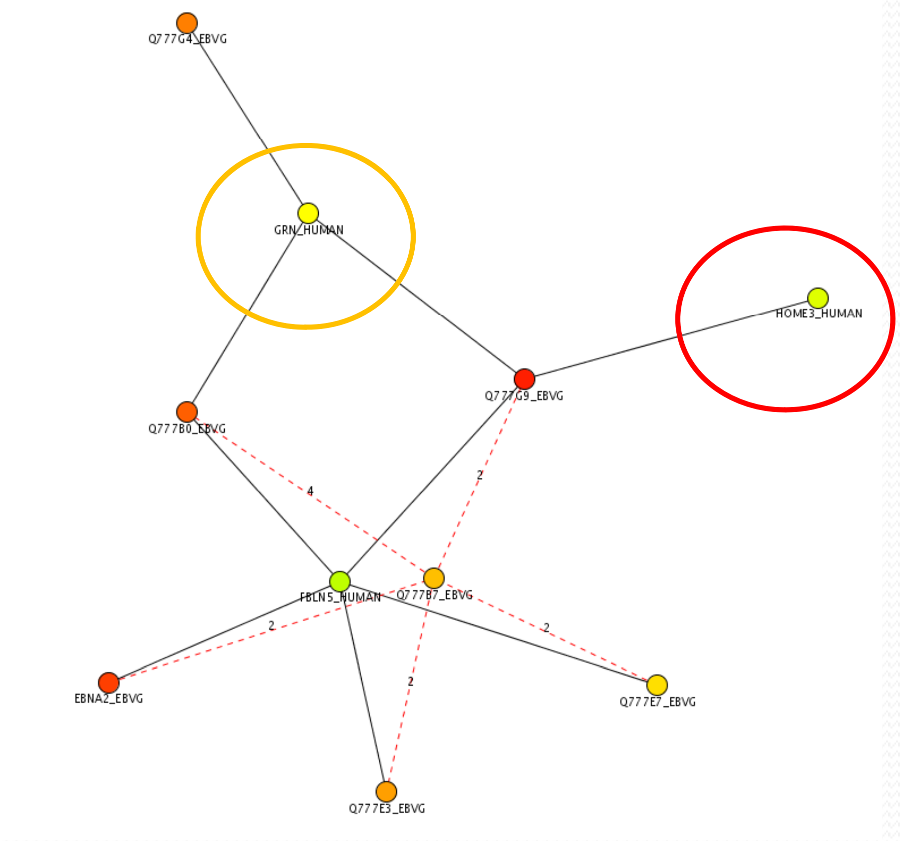
$$T = \sum_{i=1}^{79} 1 \text{ if } \log_2 \frac{eA_i}{\overline{eA}} \geq 1 \text{ and } \log_2 \frac{eB_i}{\overline{eB}} \geq 1$$

- Tissue specificity score (T)
- Sub-cellular localization score (L)
- Cell-cycle stage score (P)

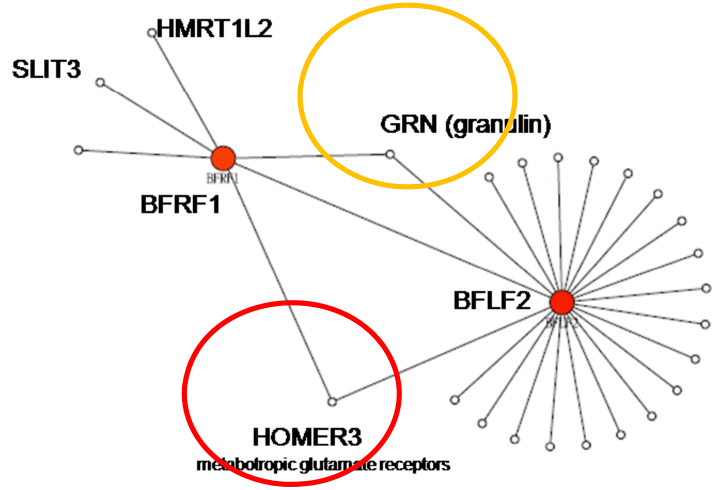
$$CS = w_I * \frac{I}{\overline{I_R}} + w_D * \frac{D}{\overline{D_K}} + w_T * \frac{T}{\overline{T_K}} + w_L * \frac{L}{\overline{L_K}} + w_P * \frac{P}{\overline{P_K}}$$

Interactome among Pathogens and Host

There are 148 nodes and 172 edges in your network. The clustering coeffcient of this network is 0 , and the average path length of this network is 3.24812 .



BFLF2 and BFRF1, Identified from EBV (intra) and Human protein interaction network (PANS, 2007)



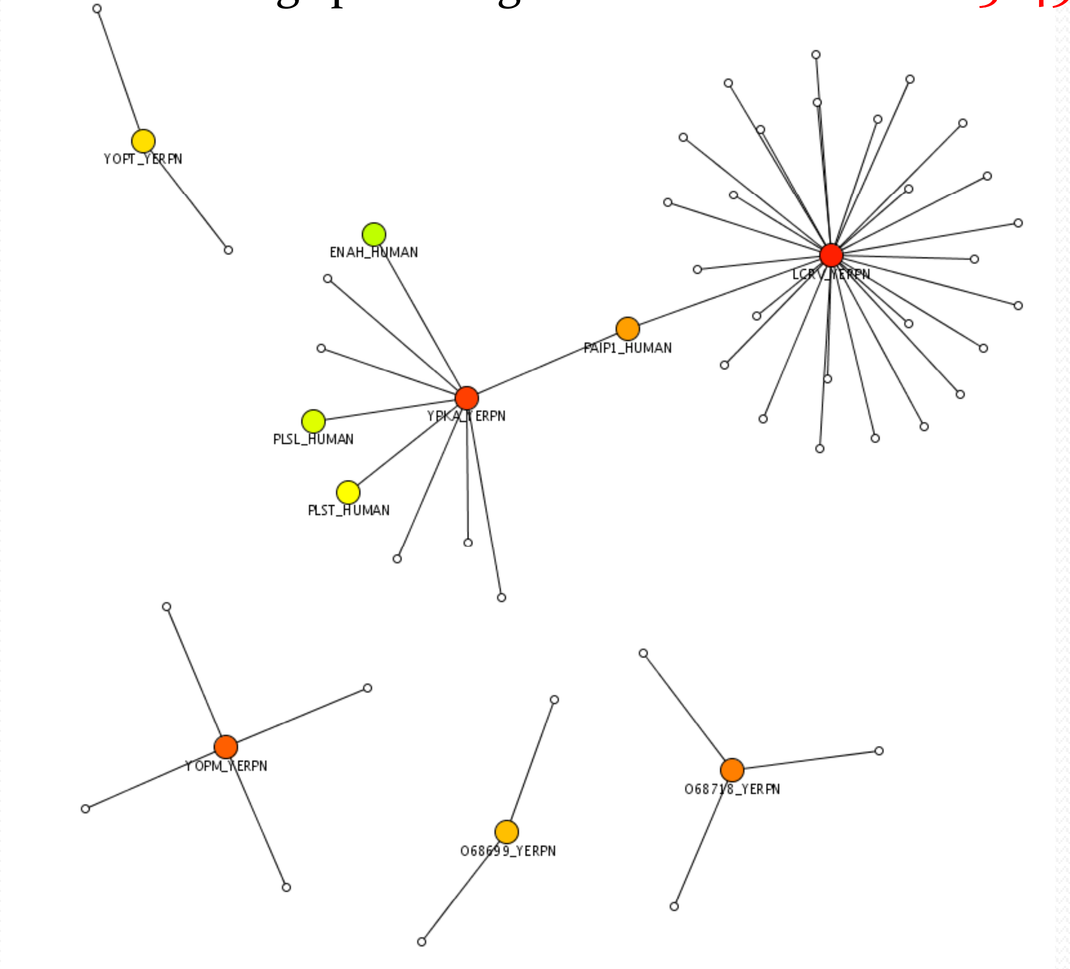
Supported citation: BFRF1 of Epstein-Barr virus is essential for efficient primary viral envelopment and egress. J Virol. 2005 Mar;79(6):3703-12.

Source: EBV and Human, Dyer *et al.*, 2008

Source: Our own with Vidal *et al.*, 2007

Interactome of Yersinia pestis and Human host

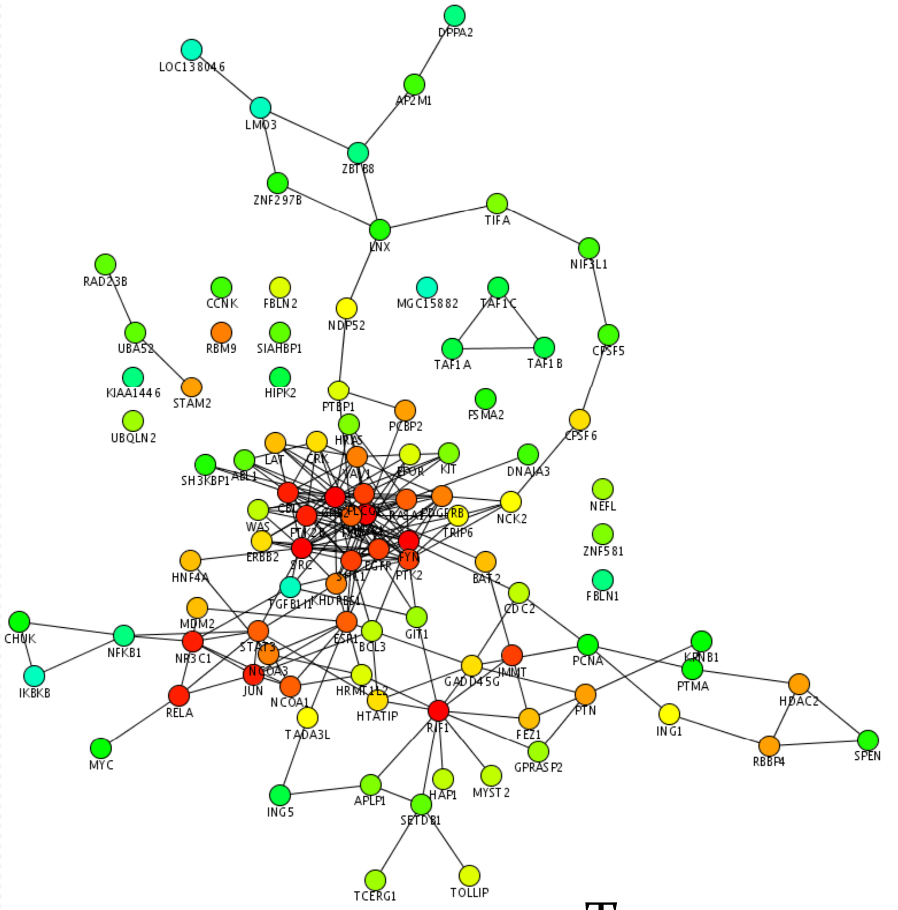
There are 56 nodes and 49 edges in your network. The clustering coeffcient of this network is 0 , and the average path length of this network is 1.15649 .



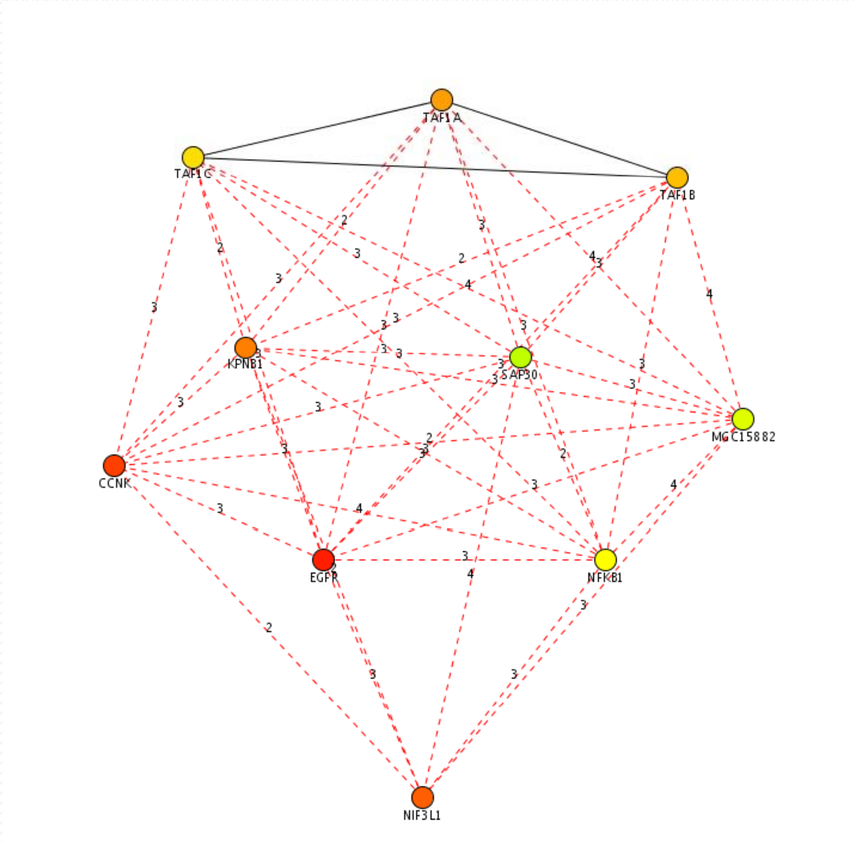
http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_30_02_36_24

Protein-protein Interactions For The Ataxia Network

There are 3607 nodes and 6972 edges in your network. The clustering coefficient of this network is 0.0570338 , and the average path length of this network is 4.18696 .



Top 100

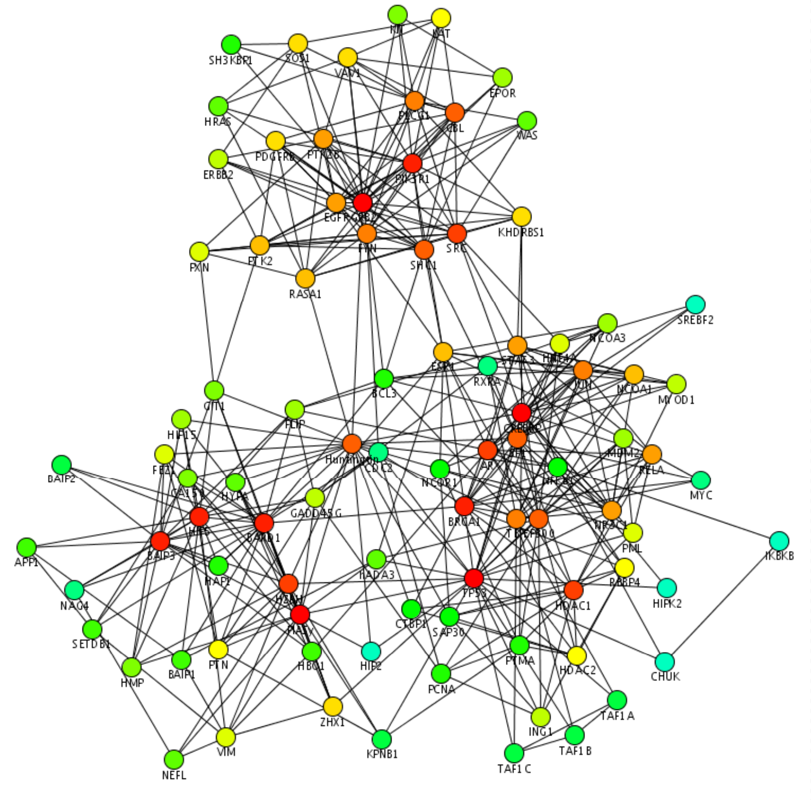


Top 10

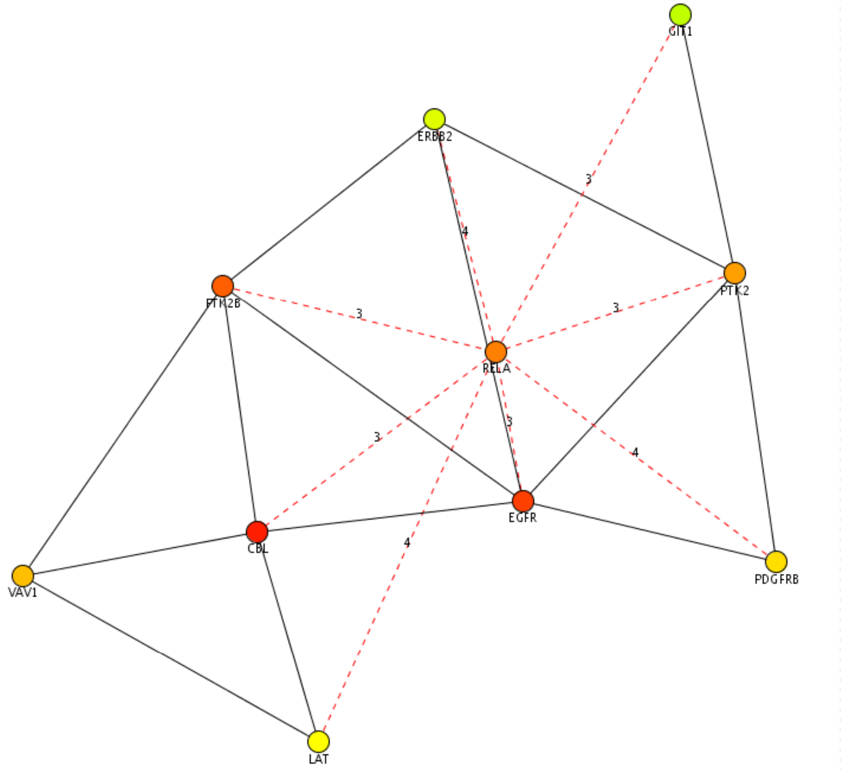
http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_20_00_44_35

A Protein Interaction Network Links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's Disease

There are 182 nodes and 592 edges in your network. The clustering coeffcient of this network is 0.23954 , and the average path length of this network is 2.85459 .



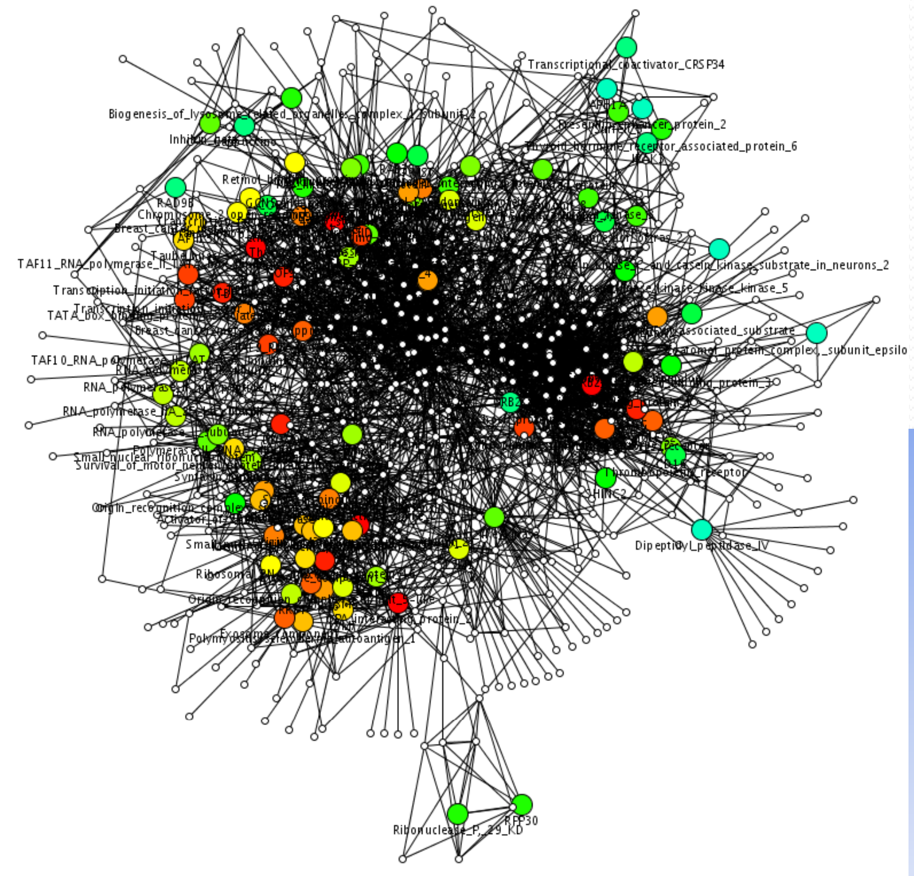
Top 90



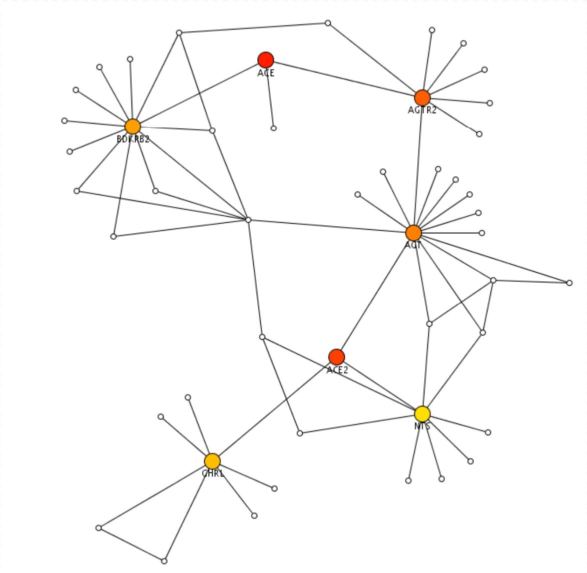
Top 10

http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_16_23_12_26

Extract Sub-network with Targeted Genes


Transcriptional_coactivator_CRSP34
Biogenesis_of_lysosome-related_organelles_complex_1_subunit_2
Thyroid_hormone_receptor_associated_protein_6
RAD9B
TAF11_RNA_polymerase_II
TAF10_RNA_polymerase
Dipeptidyl_peptidase_IV
Ribonuclease_P_29_KD
RFP30

Targeted Genes
ACE
ACE2
AGTR2
AGT
BDKRB2
GHRL
NTS
AGTR1
ARRB2

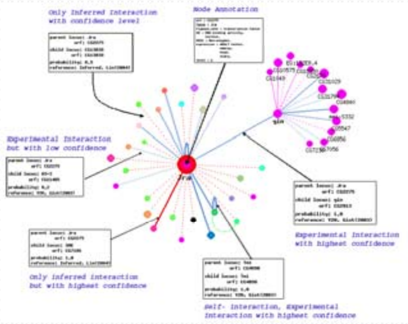


Relationships among these proteins

Human Interactome, HPRD, 2007/9/01

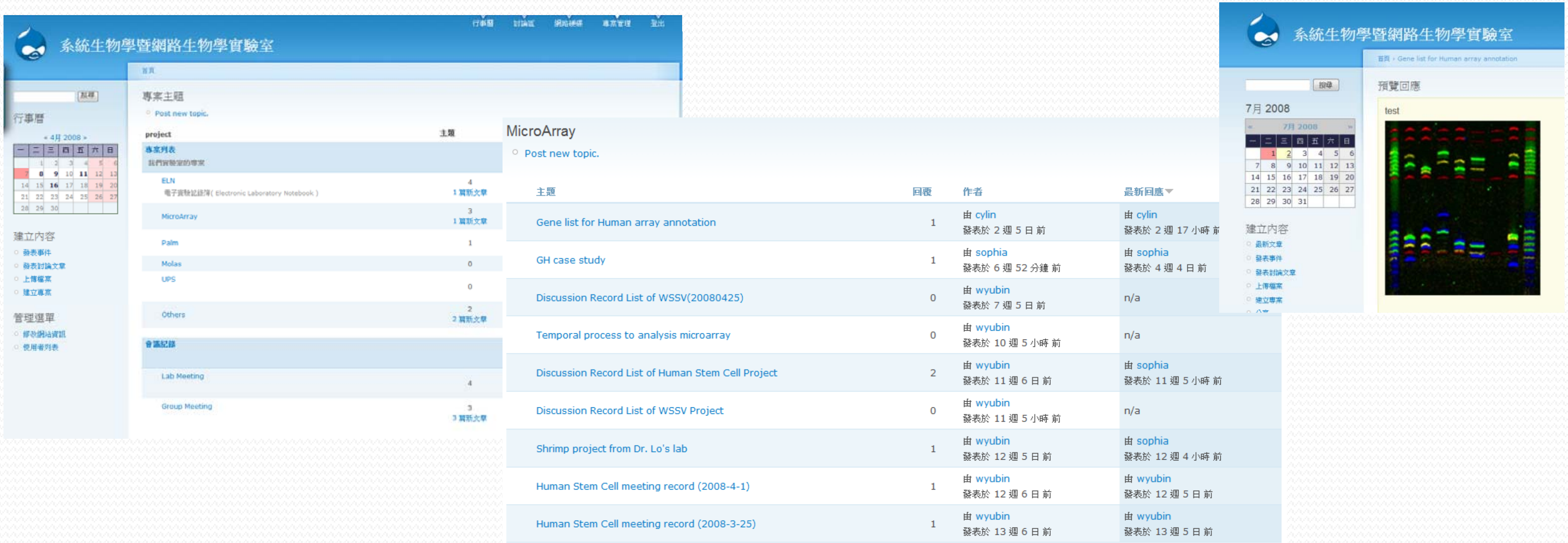
Ongoing Projects

- *Human Stem cell* research and *regenerative medicine* for stemness on expression profile and TF regulatory network (Collaborated with GRC, Academia Sinica)
- Protein interactions in the approaches of network analysis and systems biology for human and several model organisms on various spatiotemporal scenarios (granted by NRPGM)
- Electronic Lab Notebook (ELN)


Only Inferred Interaction
with confidence level
Node Annotation
Experimental Interaction
but with low confidence
Experimental Interaction
with highest confidence
Only inferred interaction
but with highest confidence
Self- interaction, experimental
interaction with highest confidence

Electronic Lab Notebook (ELN)

- Digitalization of Lab notebook from text, gif, raw data, even animations with functions of full text search and security
- Two kinds of version will be provided in the end of this year.
  - For group use: Linux-based version
  - For personal use: USB-ELN, windows/ Mac-based version



系統生物學暨網路生物學實驗室

行事曆 討論區 網站連結 專案管理 登出

首頁

搜尋

行事曆

« 4月 2008 »

一	二	三	四	五	六	日
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

建立內容
- 發表事件
- 發表討論文章
- 上傳檔案
- 建立專案

管理選單
- 修改網站資訊
- 使用者列表

專案主題
- Post new topic.

project	主題
專案列表 我們實驗室的專案	
ELN 電子實驗記錄簿(Electronic Laboratory Notebook)	4 1 篇新文章
MicroArray	3 1 篇新文章
Palm	1
Molas	0
UPS	0
Others	2 2 篇新文章
會議紀錄	
Lab Meeting	4
Group Meeting	3 3 篇新文章

MicroArray
- Post new topic.

主題	回覆	作者	最新回應
Gene list for Human array annotation	1	由 cylin 發表於 2 週 5 日 前	由 cylin 發表於 2 週 17 小時 前
GH case study	1	由 sophia 發表於 6 週 52 分鐘 前	由 sophia 發表於 4 週 4 日 前
Discussion Record List of WSSV(20080425)	0	由 wyubin 發表於 7 週 5 日 前	n/a
Temporal process to analysis microarray	0	由 wyubin 發表於 10 週 5 小時 前	n/a
Discussion Record List of Human Stem Cell Project	2	由 wyubin 發表於 11 週 6 日 前	由 sophia 發表於 11 週 5 小時 前
Discussion Record List of WSSV Project	0	由 wyubin 發表於 11 週 5 小時 前	n/a
Shrimp project from Dr. Lo's lab	1	由 wyubin 發表於 12 週 5 日 前	由 sophia 發表於 12 週 4 小時 前
Human Stem Cell meeting record (2008-4-1)	1	由 wyubin 發表於 12 週 6 日 前	由 wyubin 發表於 12 週 5 日 前
Human Stem Cell meeting record (2008-3-25)	1	由 wyubin 發表於 13 週 6 日 前	由 wyubin 發表於 13 週 5 日 前

系統生物學暨網路生物學實驗室

首頁 › Gene list for Human array annotation

搜尋

7月 2008

« 7月 2008 »

一	二	三	四	五	六	日
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

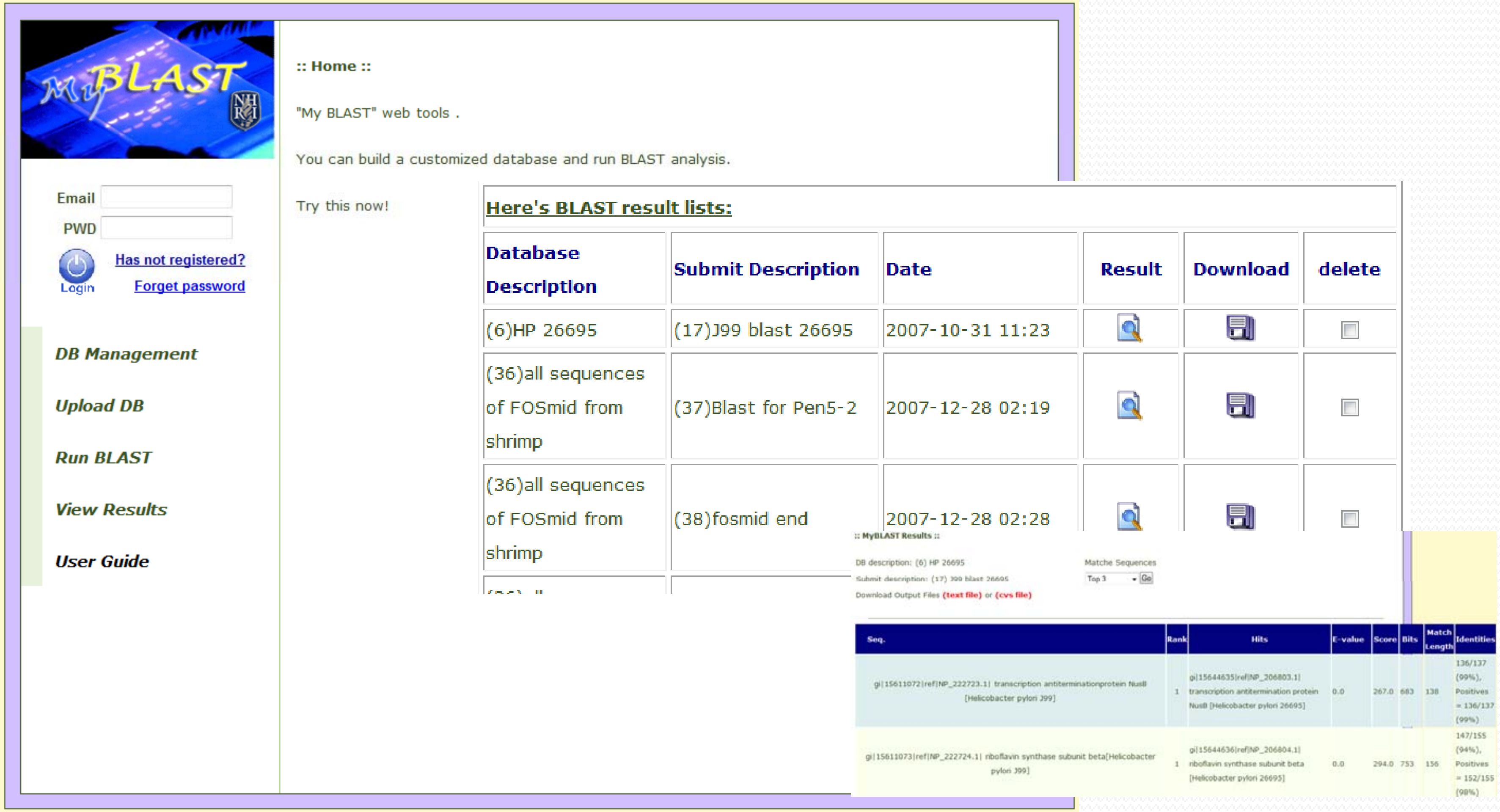
建立內容
- 最新文章
- 發表事件
- 發表討論文章
- 上傳檔案
- 建立專案

預覽回應

test

MyBLAST (Customized BLAST Framework)

http://mybioweb.nhri.org.tw/myblast



:: Home ::

"My BLAST" web tools .

You can build a customized database and run BLAST analysis.

Try this now!

Here's BLAST result lists:

Database Description	Submit Description	Date	Result	Download	delete
(6)HP 26695	(17)J99 blast 26695	2007-10-31 11:23	🔍	💾	□
(36)all sequences of FOSmid from shrimp	(37)Blast for Pen5-2	2007-12-28 02:19	🔍	💾	□
(36)all sequences of FOSmid from shrimp	(38)fosmid end	2007-12-28 02:28	🔍	💾	□

:: MyBLAST Results ::

DB description: (6) HP 26695

Submit description: (17) J99 blast 26695

Download Output Files (text file) or (cvs file)

Matche Sequences: Top 3 Go

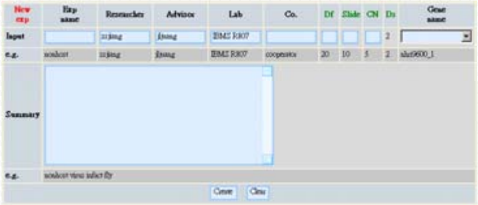
Seq.	Rank	Hits	E-value	Score	Bits	Match Length	Identities
gi|15611072|ref|NP_222723.1| transcription antiterminationprotein NusB [Helicobacter pylori J99]	1	gi|15644635|ref|NP_206803.1| transcription antitermination protein NusB [Helicobacter pylori 26695]	0.0	267.0	683	138	136/137 (99%), Positives = 136/137 (99%)
gi|15611073|ref|NP_222724.1| riboflavin synthase subunit beta[Helicobacter pylori J99]	1	gi|15644636|ref|NP_206804.1| riboflavin synthase subunit beta [Helicobacter pylori 26695]	0.0	294.0	753	156	147/155 (94%), Positives = 152/155 (98%)

MOLAS

- MicroArray On Line Analysis System (MOLAS): a web-based customizable bioinformatics package designed for manager and analyze massive array data

Web interface

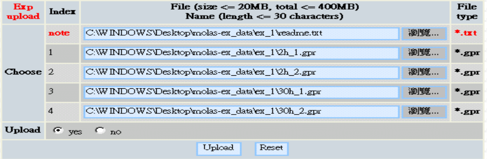
New Experiment Design



New exp	Exp name	Researcher	Advisor	Lab	Co.	Df	Slide	CN	Ds	Gene name
Input		zzjiang	jlsung	IBMS R807					2	
e.g.	[illegible]	zzjiang	jlsung	IBMS R807	cooperator	20	10	5	2	ahr9600_1
Summary										
e.g.	[illegible] virus infect fly									

Create Clear

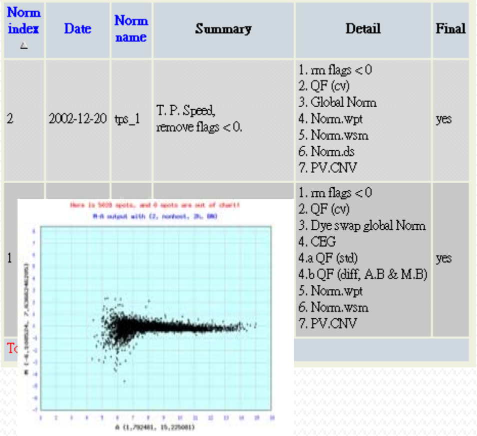
Upload raw data



Exp upload	Index	File (size <= 20MB, total <= 400MB) Name (length <= 30 characters)	File type
Choose	note	C:\WINDOWS\Desktop\molas-ex_data\ex_1\readme.txt 瀏覽...	*.txt
	1	C:\WINDOWS\Desktop\molas-ex_data\ex_1\2h_1.gpr 瀏覽...	*.gpr
	2	C:\WINDOWS\Desktop\molas-ex_data\ex_1\2h_2.gpr 瀏覽...	*.gpr
	3	C:\WINDOWS\Desktop\molas-ex_data\ex_1\30h_1.gpr 瀏覽...	*.gpr
	4	C:\WINDOWS\Desktop\molas-ex_data\ex_1\30h_2.gpr 瀏覽...	*.gpr
Upload	● yes ○ no		

Upload Reset

Exp – Experiment Data



Norm index	Date	Norm name	Summary	Detail	Final
2	2002-12-20	tps_1	T. P. Speed, remove flags < 0.	1. rm flags < 0 2. QF (cv) 3. Global Norm 4. Norm.wpt 5. Norm.wsm 6. Norm.ds 7. PV.CNV	yes
1				1. rm flags < 0 2. QF (cv) 3. Dye swap global Norm 4. CEG 4.a QF (std) 4.b QF (diff, A.B & M.B) 5. Norm.wpt 6. Norm.wsm 7. PV.CNV	yes

Here is 5020 spots, and 0 spots are out of chart!
M-A output with (2, nonhost, 2h, BN)
A (1.702481, 15.225081)

Data Normalization

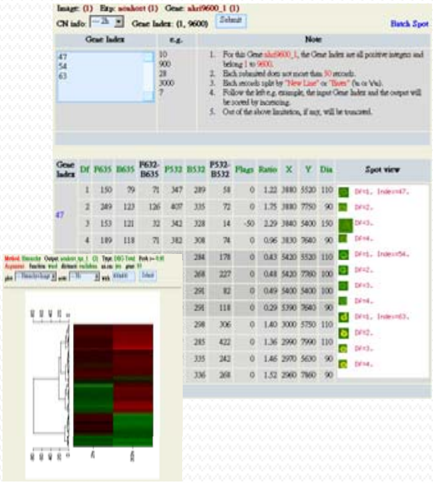


Image: (1) Exp: nonhost (1) Gene: ahr9600_1 (1)

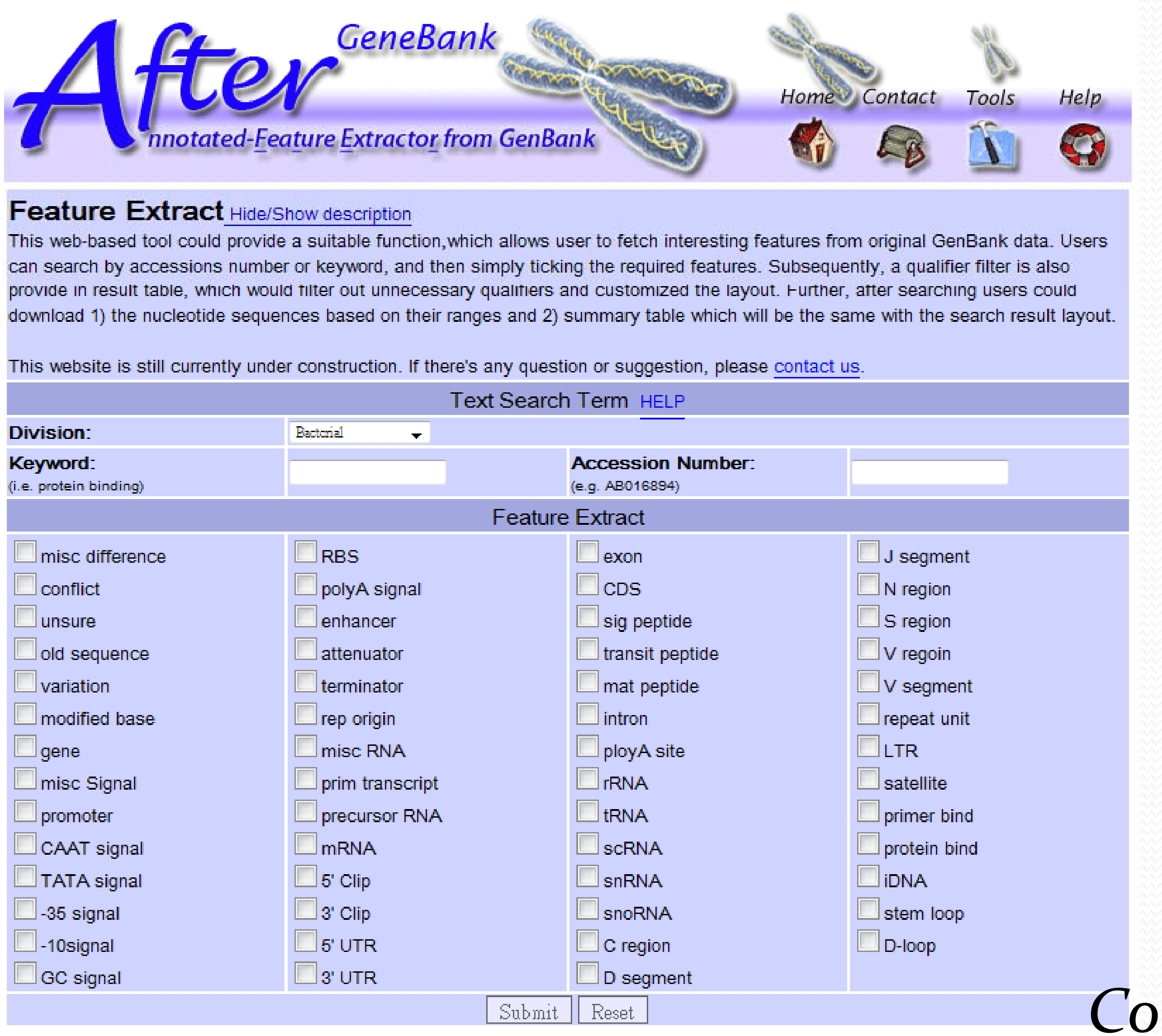
CN info: --2h Gene Index: (1, 9600) Submit Batch Spot

Gene Index	e.g.	Note
47 54 63	10 900 28 3000 7	1. For this Gene ahr9600_1, the Gene Index are all positive integers and belong 1 to 9600. 2. Each submitted does not more than 50 records. 3. Each records split by "New Line" or "Enter" (\n or \r\n). 4. Follow the left e.g. example, the input Gene Index and the output will be sorted by increasing. 5. Out of the above limitation, if any, will be truncated.

Gene Index	Df	P635	B635	P632-B635	P532	B532	P532-B532	Flags	Ratio	X	Y	Dia	Spot view
47	1	150	79	71	347	289	58	0	1.22	3880	5520	110	Df=1, Index=47,
	2	249	123	126	407	335	72	0	1.75	3880	7750	90	Df=2,
	3	153	121	32	342	328	14	-50	2.29	3840	5400	150	Df=3,
	4	189	118	71	382	308	74	0	0.96	3830	7640	90	Df=4,
						284	178	0	0.43	5420	5520	110	Df=1, Index=54,
						268	227	0	0.48	5420	7780	100	Df=2,
						291	82	0	0.49	5400	5400	100	Df=3,
						291	118	0	0.29	5390	7640	90	Df=4,
						298	306	0	1.40	3000	5750	110	Df=1, Index=63,
						285	422	0	1.36	2990	7990	110	Df=2,
						335	242	0	1.46	2970	5630	90	Df=3,
						336	268	0	1.52	2960	7860	90	Df=4,

Experiment Result and Analysis

Annotated-Feature Extractor from GenBank



Feature Extract Hide/Show description

This web-based tool could provide a suitable function,which allows user to fetch interesting features from original GenBank data. Users can search by accessions number or keyword, and then simply ticking the required features. Subsequently, a qualifier filter is also provide in result table, which would filter out unnecessary qualifiers and customized the layout. Further, after searching users could download 1) the nucleotide sequences based on their ranges and 2) summary table which will be the same with the search result layout.

This website is still currently under construction. If there's any question or suggestion, please contact us.

Text Search Term HELP

Division:	Bacterial		
Keyword: (i.e. protein binding)		**Accession Number:** (e.g. AB016894)	

Feature Extract

☐ misc difference
☐ conflict
☐ unsure
☐ old sequence
☐ variation
☐ modified base
☐ gene
☐ misc Signal
☐ promoter
☐ CAAT signal
☐ TATA signal
☐ -35 signal
☐ -10signal
☐ GC signal

☐ RBS
☐ polyA signal
☐ enhancer
☐ attenuator
☐ terminator
☐ rep origin
☐ misc RNA
☐ prim transcript
☐ precursor RNA
☐ mRNA
☐ 5' Clip
☐ 3' Clip
☐ 5' UTR
☐ 3' UTR

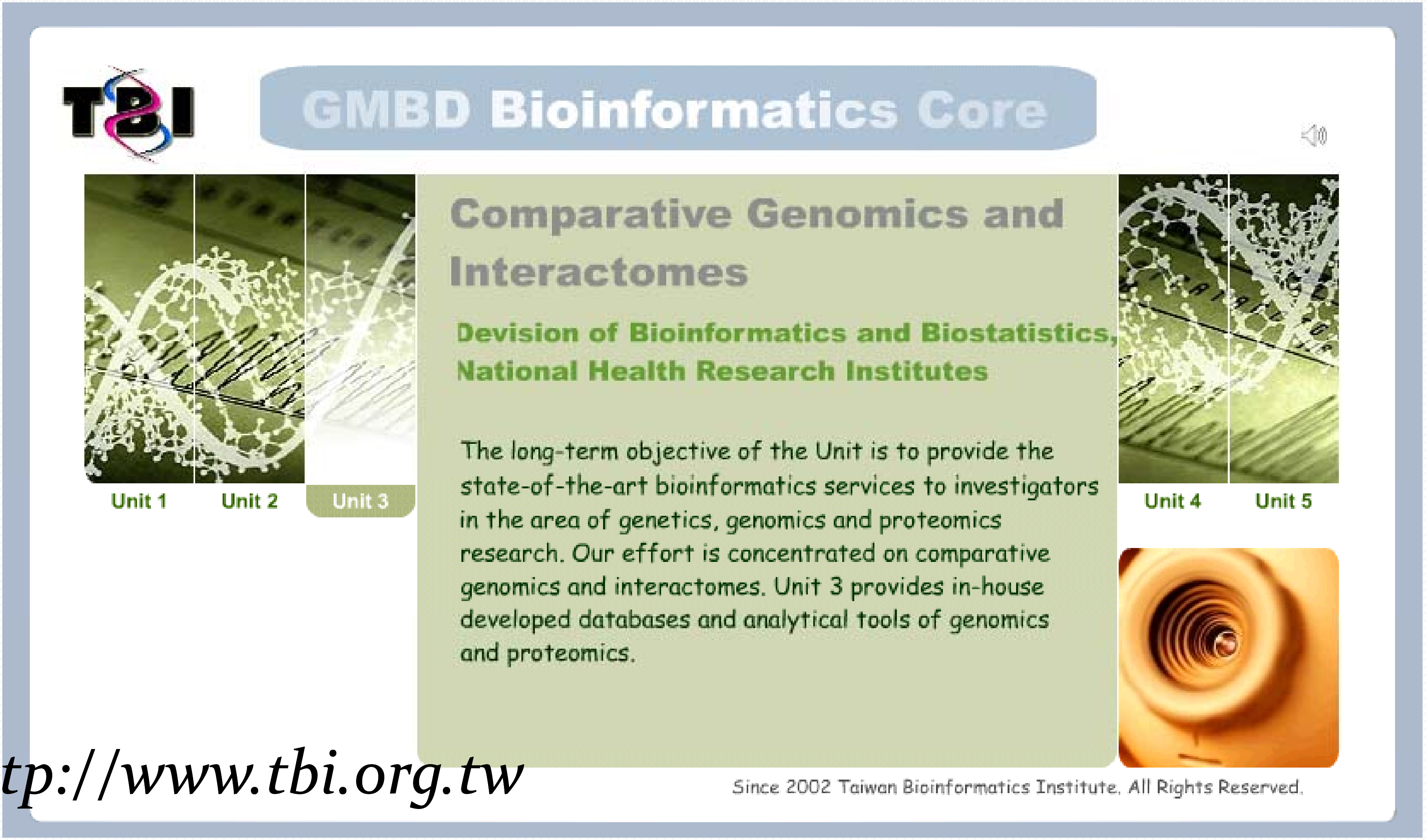
☐ exon
☐ CDS
☐ sig peptide
☐ transit peptide
☐ mat peptide
☐ intron
☐ ployA site
☐ rRNA
☐ tRNA
☐ scRNA
☐ snRNA
☐ snoRNA
☐ C region
☐ D segment

☐ J segment
☐ N region
☐ S region
☐ V regoin
☐ V segment
☐ repeat unit
☐ LTR
☐ satellite
☐ primer bind
☐ protein bind
☐ iDNA
☐ stem loop
☐ D-loop

Submit Reset

Coming Soon

Bioinformatics Core for Genomic Medicine and Biotechnology Development



GMBD Bioinformatics Core

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

Selected Publications (2006 - 8)

1) **Lin, C. Y.*,** Chin, C. H., Wu, H. H., Chen, S. H., Ho, C. W.,* Ko, M. T.*, "Hubba: Hub Objects Analyzer : A Framework of Interactome Hubs Identification for Network Biology," Nucleic Acids Res., volume 36, number 2008 Web application Issue, July 2008, ***Nucleic Acids Research*** Advance Access published online on May 24, 2008 (http://hub.iis.sinica.edu.tw) (SCI/6.945) .

2) Chen, S.H., Lo, C.Z., Tsai, M. C., Hsiung C.A., **Lin, C.Y***., **2008**. "Unique Probe Selector (UPS): A Comprehensive Web Service for Probe Design and Oligo Nucleotide Arrays," To Appear in ***BMC Bioinformatics***, (URL: http://array.iis.sinica.edu.tw/ups) (SCI/3.49) .

3) Huang, T. W., **Lin, C. Y***., Kao, C. Y. **2007**. Reconstruction of Human Protein Interolog Network using Evolutionary Conserved Network. ***BMC Bioinformatics***. 8:152 (SCI/3.49) .

4) **Lin, C.Y. ***, Chen S. H., Cho C. S., Chen C. L., Lin F. K., Lin C. H., Chen P. Y., Lo C. Z., and Hsiung C.A., **2006**, "Fly-DPI: Database of Protein Interactomes for *D. melanogaster* in the Approach of Systems Biology.," ***BMC Bioinformatics***, 7(5):S18, (SCI/3.49) (URL:http://flydpi.nhri.org.tw)

5) Jiang S. S., Chang I. S., Huang L. W., Chen P. C., Wen C. C., Liu S. C., Chien L. C., **Lin C. Y.**, Hsiung C. A., Juang J. L., **2006**"Temporal Transcription Program of Recombinant *Autographa californica* Multiple Nucleopolyhedrosis Virus.," ***J. Virol***., 80: 8989-8999. (SCI/ 5.178)

6) Wen, C. C., Wu, Y. J., Huang, Y. H., Chen, W. C., Liu, S. C., Jiang, S. S., Juang, J. L., **Lin, C. Y.**, Fang, W. T., Hsiung, C. A., Chang, I. S. **2006**. A Bayes Regression Approach to Array-CGH Data. ***Statistical Applications in Genetics and Molecular Biology***. 5(1): art3, (http://www.bepress.com/sagmb/vol5/iss1/art3/), (Medline Index)

7) Chang, C. C., Lin, H. C., Lin, I. P., Chang, T. Y., Chen, H. H., Chen, W. H. Cheng, C. H., **Lin, C. Y.**, Liu, S. M. Chang, C. C. Chaw, S. M. **2006**. The Chloroplast Genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative Analysis of Evolutionary Rate with That of Grasses and Its Phylogenetic Implications. ***Mol. Bio. Evol***. 23: 279 - 291 (SCI/ 6.355)

8) Pan W. H., Lynn K. S., Chen C. H., Wu Y. L., Lin C. Y., Chang H.Y. **2006**. Using endophenotypes for pathway clusters to map complex disease genes. ***Gen. Epi.*** 30(2): 143-154. (SCI/5.42)

Acknowledgement

國家衛生研究院
National Health Research Institutes

Chia-Ling Chen
Fan-Kai Lin
Chieh-Hwa Lin
Chia-Ming Chang
Chi-Shiang Cho
Chen-Zen Lo
Yung-Shiang Hwang
Pao-Yang Chen
Ming-Hsin Tasi
Char-Lin Pan
Chao A. Hsiung

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

Hsin-Hung Wu
Daniel, Sheng-Yao, Su
Pan-Han, Kuo
Tengi, Huang
Yu-Bin, Wang
Ming-Ta Ko

國立中央大學
資訊工程學系
Department of Computer Science and Information Engineering National Central University

Chia-Hau Kim
Chin-Wen Ho

臺大資訊

Chen-Yen Kao
Tao-Wei Huang

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

Thanks for your Attention