

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*

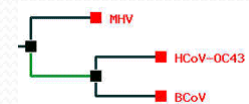
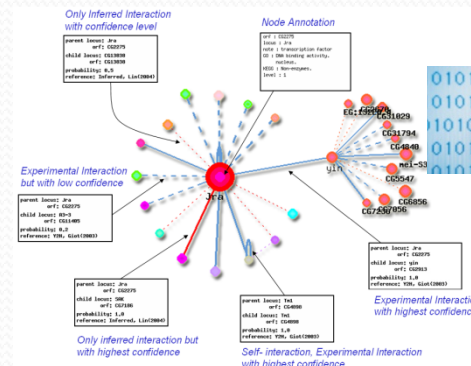
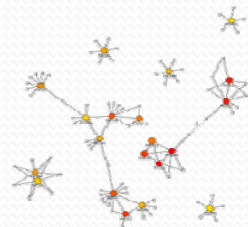
July 22, 2009



國家衛生研究院
National Health Research Institutes

Outline

- **Using IT innovations to improve the quality of life (QoL)**
 - Protein interactome
 - Network comparison and topological analysis
 - Genomics studies for high throughput research
 - Phylogenetics analysis
 - Ongoing projects



Platform Based on LAMP/ LAPP

Linux

Operation System

Apache (with OpenSSL)

Webserver

MySQL/**P**ostgreSQL

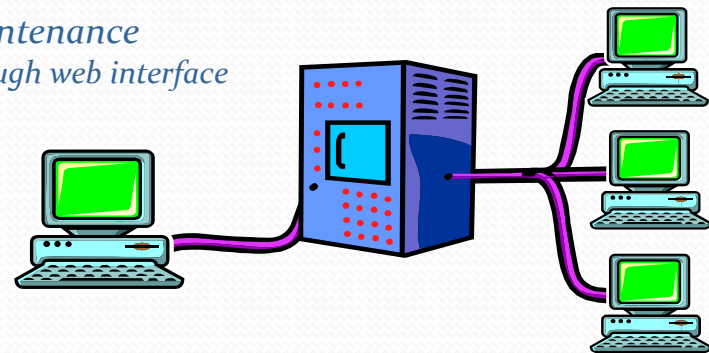
Relational Database

PHP

Server-side HTML embedded
scripting language with GD
library

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

*Maintenance
through web interface*



Web database

Linux

Apache webServer,

MySQL

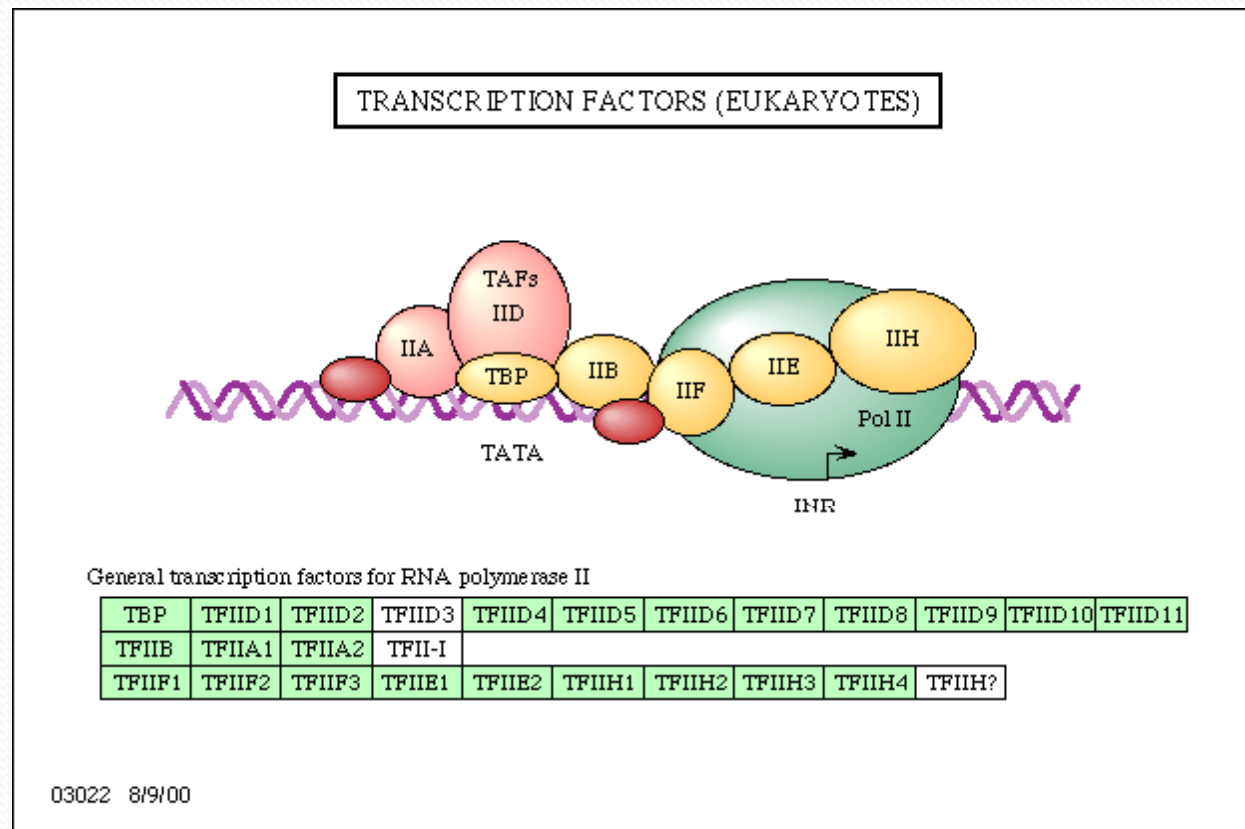
PHP control language

Interactomics: From protein to network, protein-protein interactions databases

- ✓ *hp*-DPI
- ✓ *fly*DPI
- ✓ Reconstruction of Human protein network
- ✓ Topological analysis by Hubba



Protein Interactions and Function



Graph coming from <http://www.genome.ad.jp/kegg/pathway/sce/sce03022.html>

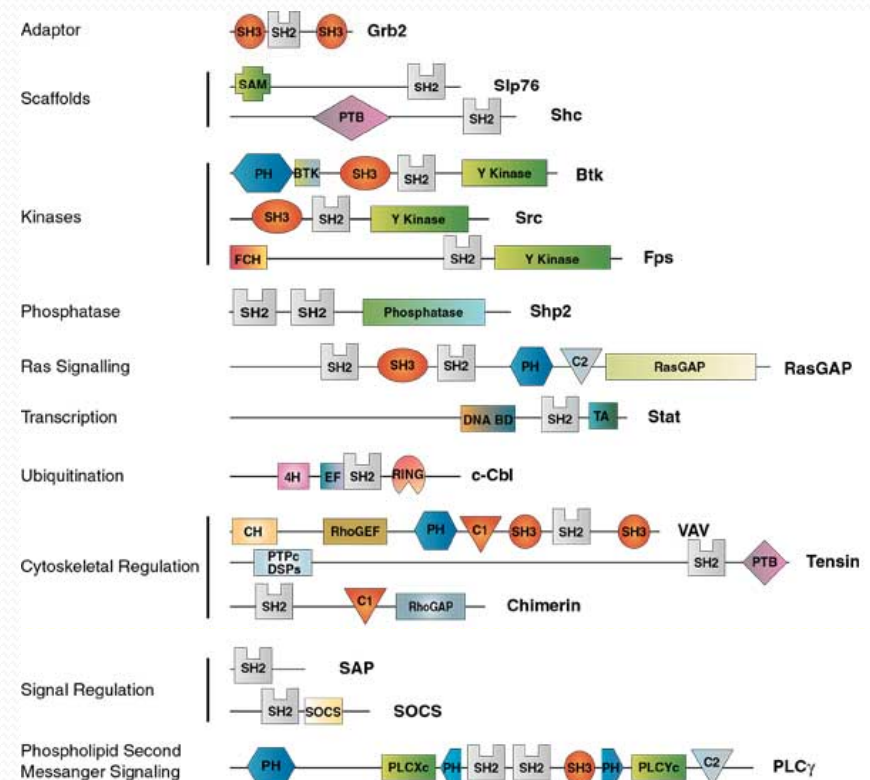


Motivations

- 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**

The Meaning of Interaction Domains

- Interaction domains are often used repeatedly in numerous different proteins to mediate a particular type of molecular recognition.
- Databases for deciphering Proteins to domain: Pfam, SMART, Procite, PRINTS, Prodom, and Interpro

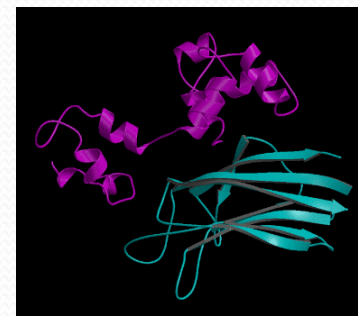


Deciphering Protein into Domains

- 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 allow us to discover novel interactions between proteins that contain domains with known binding partner.



Protein Interactions

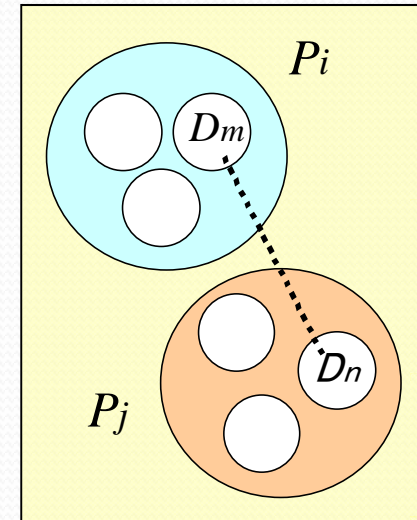


Domain Interactions

Inferred Human Protein Interactions Based on Conserved Domain Interactions

- Statistical Model

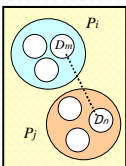
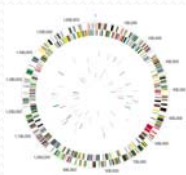
- At least one pair of domains interact together, the proteins contain these domains can be assumed to interact.
- The possibility of each domain pair is independent.
- $P_{ij} = 1$: Interaction between two protein P_i and P_j
- $D_{mn} = 1$: Interaction between two domains, D_m and D_n



$$\text{Prob}(P_{ij} = 1) = 1 - \prod_{D_{mn} \in P_{ij}} (1 - \text{Prob}(D_{mn} = 1))$$

Previous Work I:

Helicobacter pylori- Database of Protein Interactome



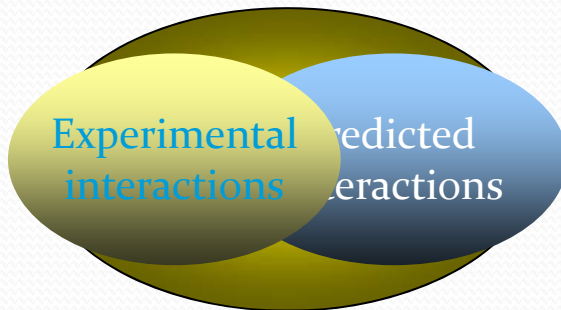
Letters to Nature

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

The protein-protein interaction map of *Helicobacter pylori*

Jean-Christophe Rain¹, Luc Selig¹, Hilde De Reuse², Véronique Battaglia¹, Céline Reverdy¹, Stéphane Simon¹, Gerlinde Lenzen¹, Fabien Petel¹, Jérôme Wojcik¹, Vincent Schächter¹, Y. Chemama¹, Agnès Labigne² and Pierre Legrain¹

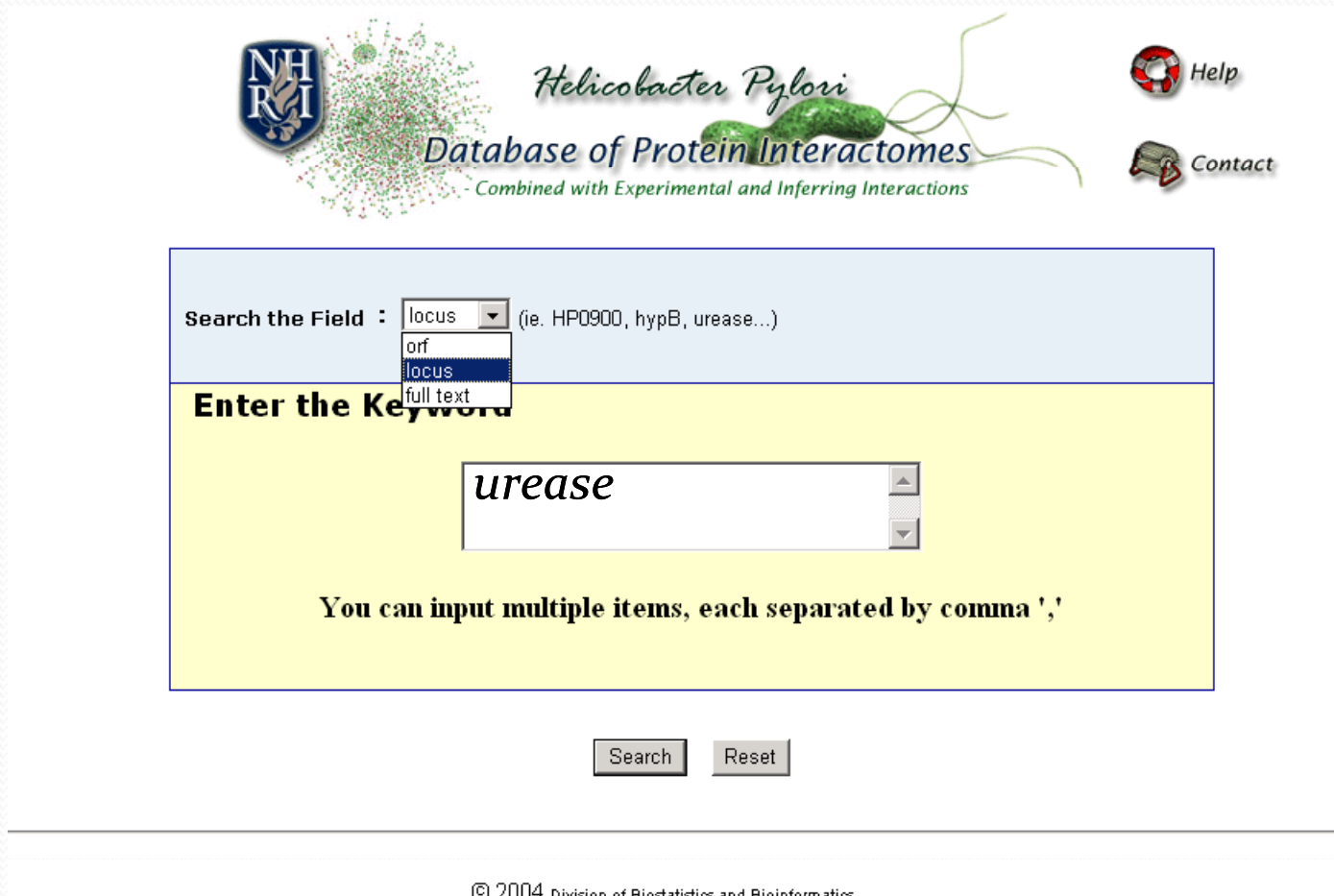
Over 1,200 interactions were identified between *H. pylori* (strain 26695) proteins, connecting 46.6% of the proteome.



© 2004 Division of Bioinformatics and Biostatistics

The network of whole proteome

Previous Work, hp-DPI



The screenshot shows the homepage of the Helicobacter Pylori Database of Protein Interactomes (hp-DPI). At the top left is the NHRI logo. The main title is "Helicobacter Pylori Database of Protein Interactomes" with a subtitle "Combined with Experimental and Inferring Interactions". There are "Help" and "Contact" links on the right. The search interface is a light blue box with a dropdown menu for "Search the Field" (options: locus, orf, locus, full text) and a text input field containing "urease". Below the input field is a note: "You can input multiple items, each separated by comma ','". At the bottom are "Search" and "Reset" buttons. The footer contains the copyright notice: "© 2004 Division of Biostatistics and Bioinformatics".

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

Enter the Keyword

urease

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

Search Reset

© 2004 Division of Biostatistics and Bioinformatics

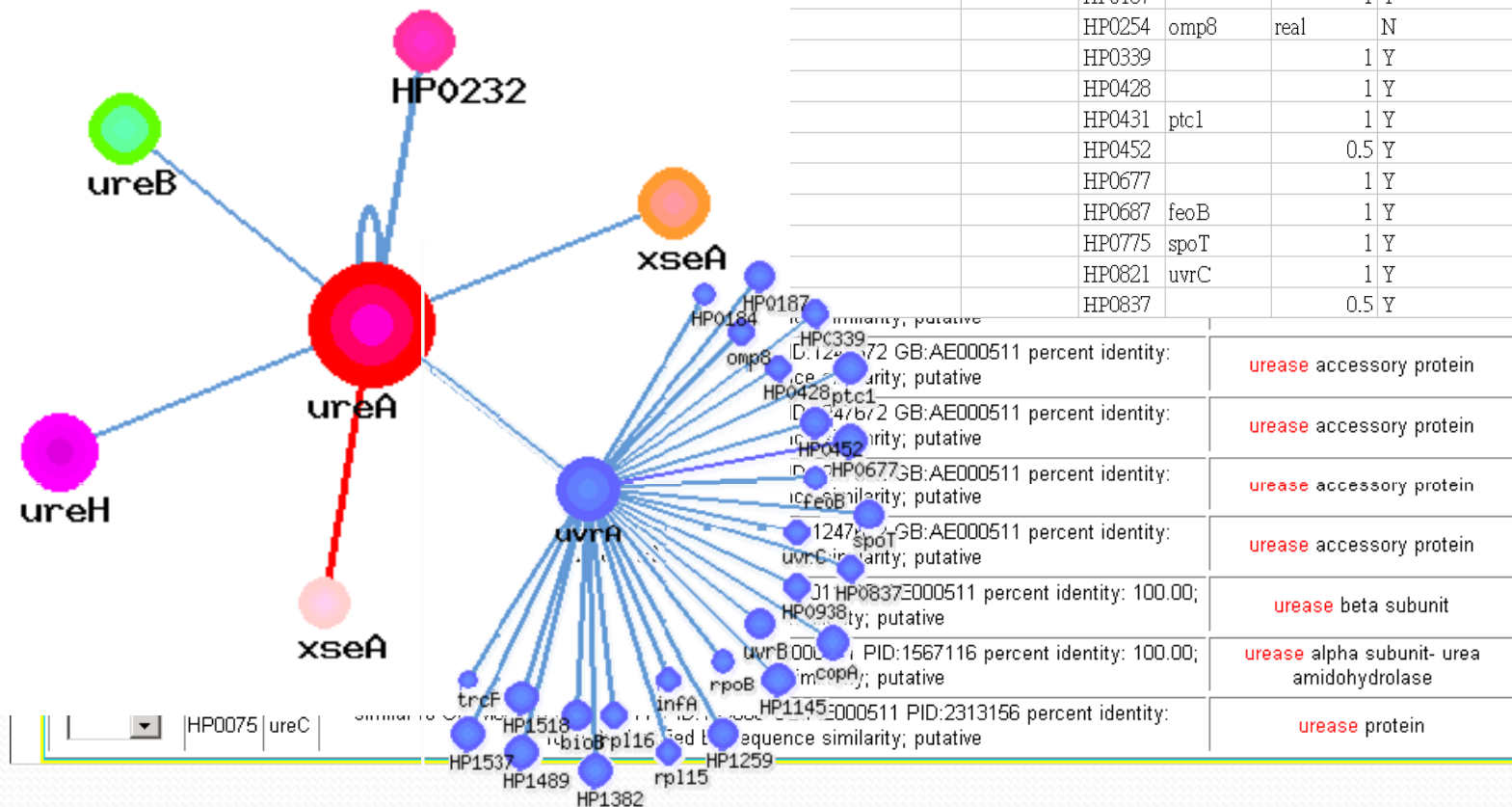
This website can be accessed at <http://dpi.nhri.org.tw/hp/>

Search Result of hp-DPI

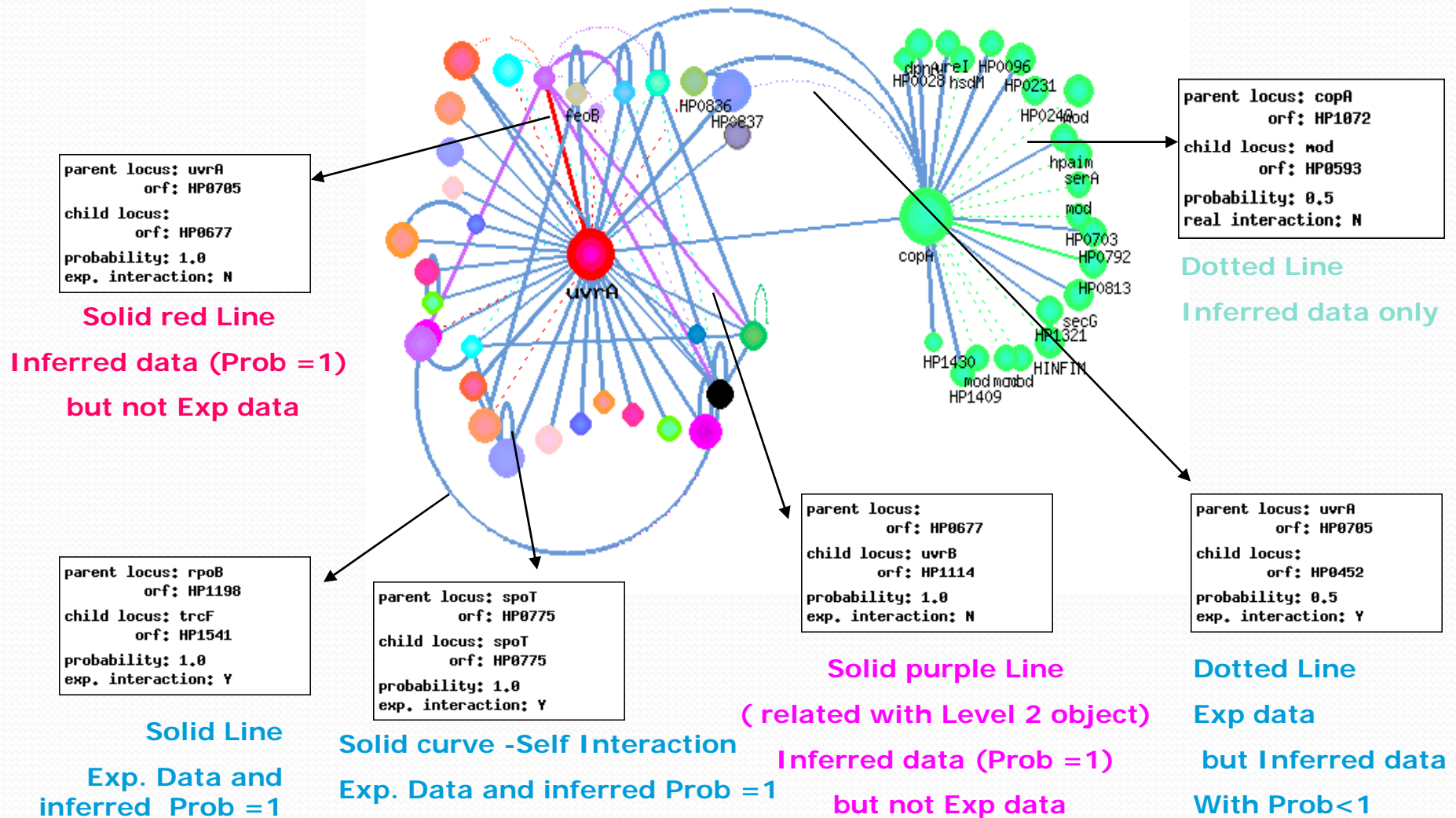
You search by: **urease**

Search type: **full text**

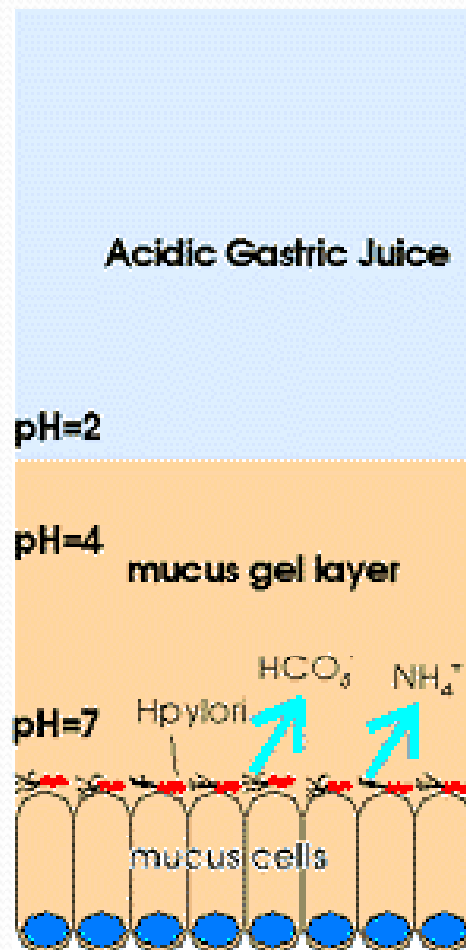
	A	B	C	D	E	F
1	Database: hp					
2	Probability range: ≥ 0.6					
3	Total: 52					
4	parent orf	parent locus	child orf	child locus	probability	real interaction
5	HP0705	uvrA	HP0073	ureA	1	Y
6			HP0184		1	Y
			HP0187		1	Y
			HP0254	omp8	real	N
			HP0339		1	Y
			HP0428		1	Y
			HP0431	ptc1	1	Y
			HP0452	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



Edges Patterns for Interaction



Discover New Research Targets with *hp*-DPI

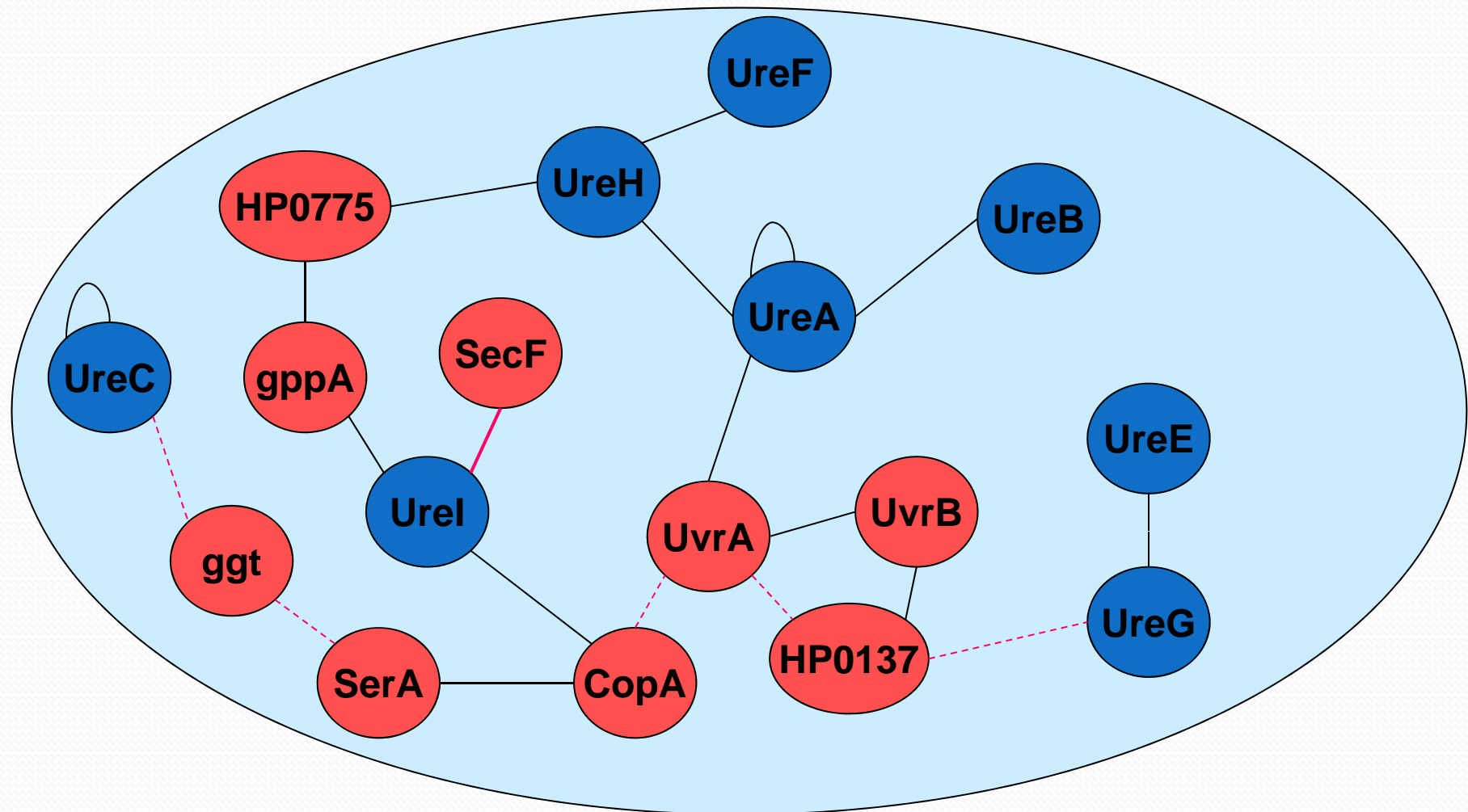


Urease

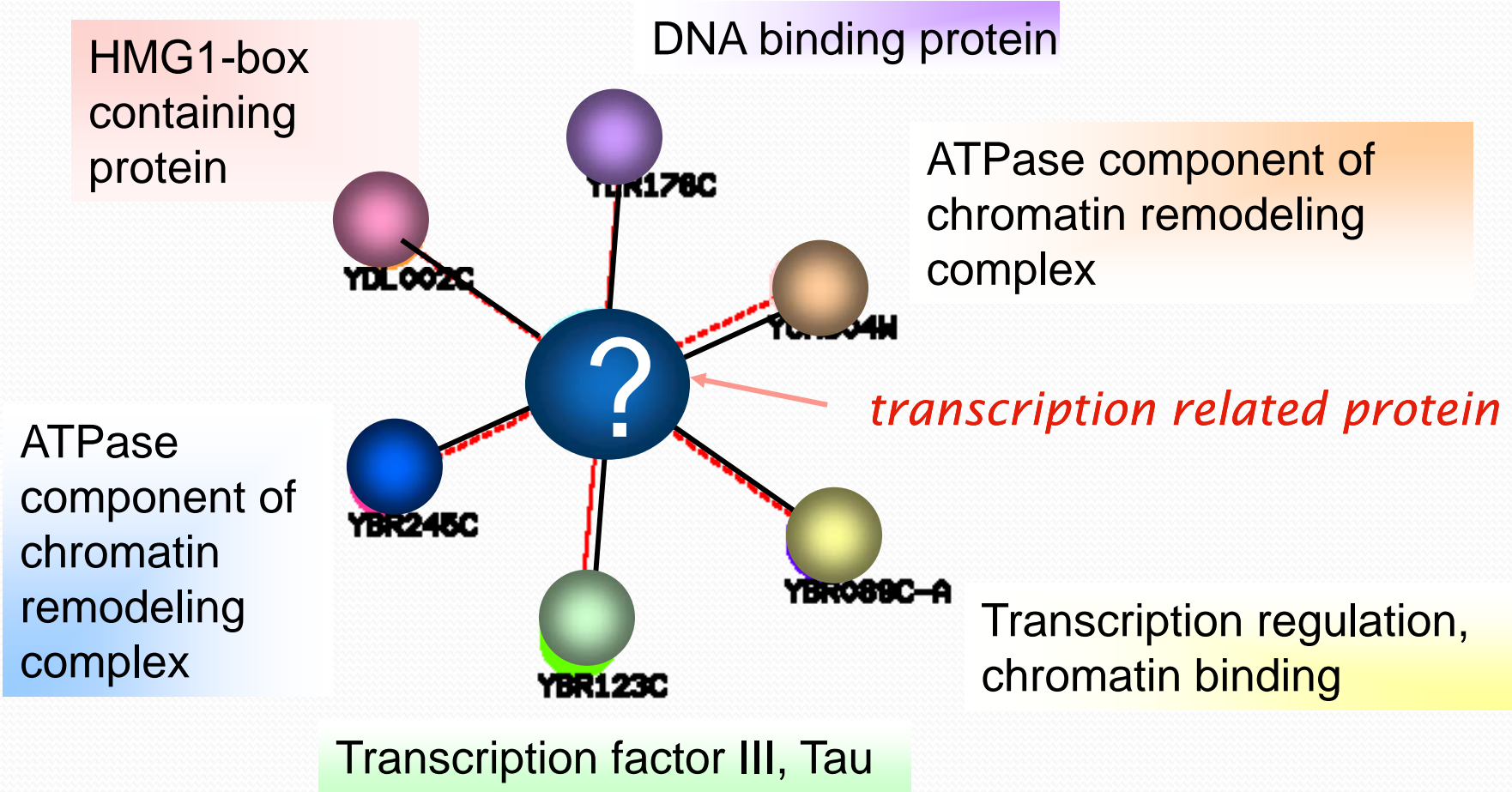


bases

Network of Urease Complex



Annotated Protein Function by Interacting Network



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

BIOINFORMATICS	
Author:	Keyword(s):
Year:	Vol: Page:

Institution: National Health Research Institutes [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

***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. Hsiung ¹

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

* To whom correspondence should be addressed.

Chung-Yen Lin, E-mail: cylin@nhri.org.tw



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

International Journal of Medical Microbiology 296 (2006) 321–324

IJMM

www.elsevier.de/ijmm

SHORT COMMUNICATION

Helicobacter pylori protein oxidation influences the colonization process

Renata Godlewska^a, Artur Dzwonek^b, Michał Mięka^b, Jerzy Ostrowski^b,
Marcin Pawłowski^c, Janusz M. Bujnicki^c, Elżbieta K. Jagusztyn-Krynicka^{a,*}

^aDepartment of Bacterial Genetics, Institute of Microbiology, Warsaw University, Miecznikowa 1, PL-02-096 Warsaw, Poland

^bDepartment of Gastroenterology, Medical Center for Postgraduate Education, Maria Skłodowska-Curie Memorial Cancer Center,

R. Godlewska et al. / International Journal of Medical Microbiology 296 (2006) 321–324

Acknowledgments

Work was supported by the Ministry of Scientific Research and Information Technology (Grants no. 4 P04 013 17 and 6 P06 K 043 21).

References

S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
D., Holland, P., Siejak, F., Koesling, J., Sabarth, N., S., Zimny-Arndt, U., Jungblut, P.R., Meyer, T.F.,

Kadokura, H., 2003. Protein disulfide bond formation in prokaryotes. *Annu. Rev. Biochem.* 72, 111–135.

Lamarque, D., Peek, R.M.J., 2003. Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 8, 21–30.

Lee, A., O'Rourke, J., De, U.M.C., Robertson, B., Daskalopoulos, G., Dixon, M.F., 1997. A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. *Gastroenterology* 112, 1386–1397.

Lin, C.Y., Chen, C.L., Cho, C.S., Wang, L.M., Chang, C.M., Chen, P.Y., Lo, C.Z., Hsiung, C.A., 2005. hp-DPI: *Helicobacter pylori* database of protein interactomes – embracing experimental and inferred interactions. *Bioinformatics* 21, 1288–1290.

Masafumi, A., Hironori, M., Masami, I., Jun-Xiong, X., Takashi, K., Masaharu, S., Toshio, S., 2004. ConPred II: a consensus prediction method for obtaining transmembrane

ular Biology, Trojdena 4, PL-02-

5



Search the Field : (in: unesse,HP0000,hydB...)

Enter the Keyword

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

Lin, C.Y., Chen, C.L., Cho, C.S., Wang, L.M., Chang, C.M., Chen, P.Y., Lo, C.Z., Hsiung, C.A., 2005. hp-DPI: Helicobacter pylori database of protein interactomes - embracing experimental and inferred interactions. *Bioinformatics* 21, 1288–1290

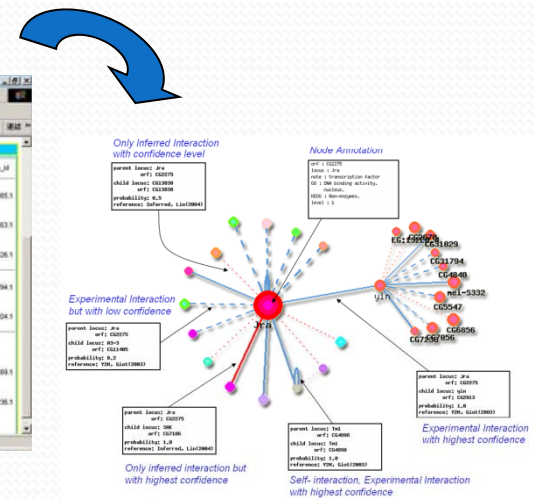
Previous Work II: Fly Database of Protein Interactomes



Query in Full Text

interaction type	gene	locus	alias	chromosome	SOD_ID	description	gene product	phenotype	genebank_id	protein_id
	YEL007C	SLA1		8	30000103	"Involved in assembly of cortical actin cytoskeleton, contains 3 SH3 domain, interacts with Baz1p"	cytoskeletal protein-binding protein	"Full mutant is viable, temperature sensitive, null mutants are synthetically lethal in combination with act1 and act1 mutants"	547695 222910 232708	AAB02985.1 CAAG0683.1 CAAG4826.1
	YOR172C	PERP	MR1216	VI	30003940	Poly(A)-binding protein-binding protein		Null mutant is viable, other mutant suppressors per1 null mutant	U88931 272963	AAG04296.1 CAAG7204.1
	YOR049C	TOR1A	EPF_A04862	XV	30009571	Involved in supporting the cooperativity and the general transcription pore and facilitating the release of pretranscripts from repress components	associates with TCM102 protein complex	"Full mutant is viable, associated with a delay of onset of prepolymerization, identification of pretranscripts, binding to receptors and the general transcription pore and destabilization of the transcription initiation complex and the general transcription pore, basal transcription factors are bound double mutants are	223015 274253	CAAG0489.1 CAAG3295.1

Search Result & Statistical Estimation



Network Visualization & popup annotation

<http://flydpi.nhri.org.tw>

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

OXFORD JOURNALS CONTACT US MY BASKET MY ACCOUNT

Nucleic Acids Research

ABOUT THIS JOURNAL CONTACT THIS JOURNAL SUBSCRIPTIONS CURRENT ISSUE ARCHIVE SEARCH

[Oxford Journals](#) > [Life Sciences](#) > [Nucleic Acids Research](#) > [Database Summary Paper 664](#)

« PREVIOUS NEXT »

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](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

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

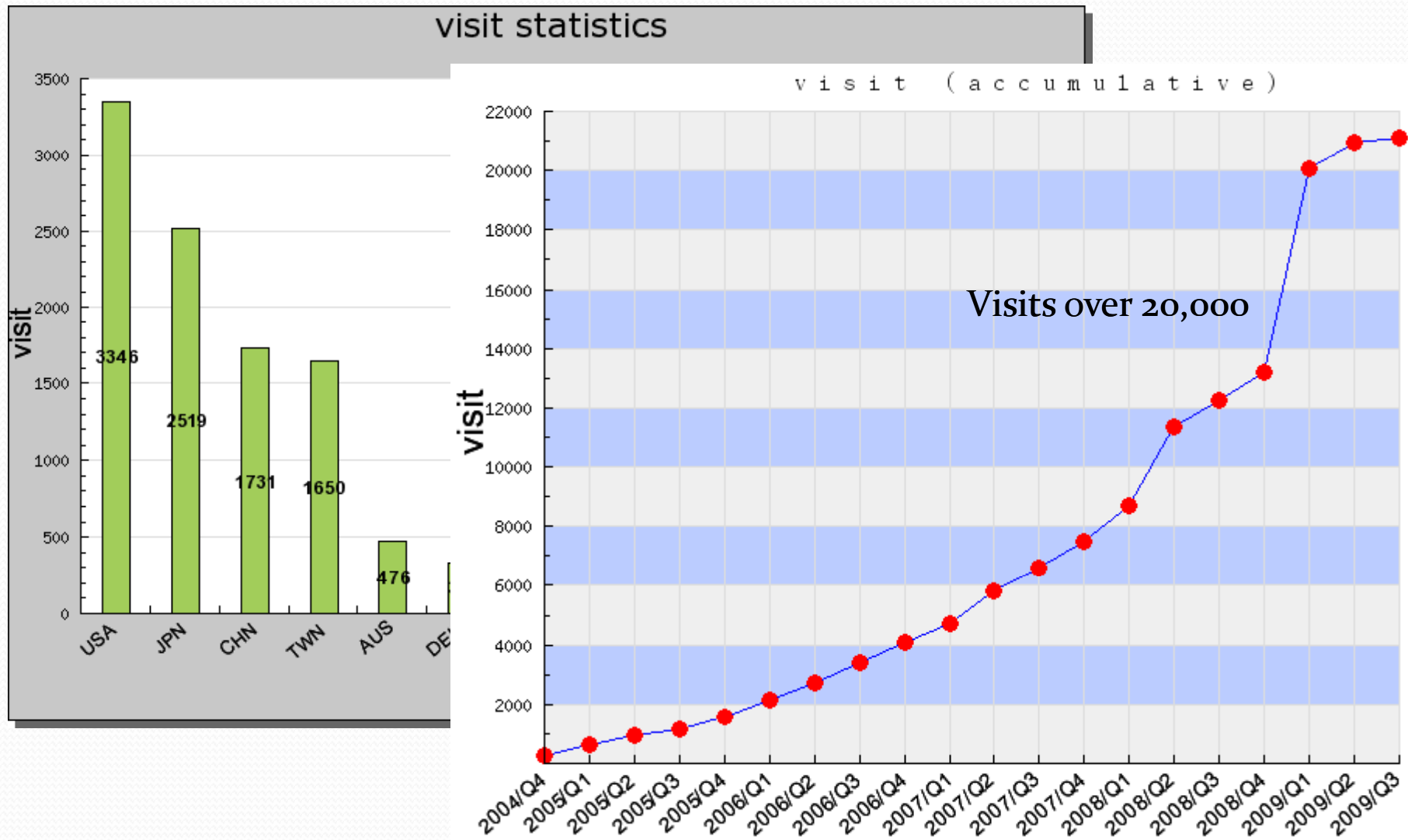
Online ISSN 1362-4962 - Print ISSN 0305-1048 Copyright © 2006 Oxford Journals

OXFORD JOURNALS
OXFORD UNIVERSITY PRESS

[Site Map](#) [Privacy Policy](#) [Frequently Asked Questions](#)

Other Oxford University Press sites:
Oxford University Press

Visit Statistics for hp-DPI from 2004/11/22 ~ 2009/7/15



Hybrid Model for Estimation

- **Maximized Likelihood Estimation (MLE)**
 - To incorporate the global consideration for whole interactomes and the experimental error ignored by local estimate, the likelihood function L is based on all potential protein interactions. O_{ij} is observed interaction. P_{ij} is predicted interaction.

$$\Pr(P_{ij} = 1) = 1 - \prod_{D_{mn} \in P_{ij}} (1 - \Pr(D_{mn} = 1)),$$

$$L = \prod \left(\Pr(o_{ij} = 1) \right)^{O_{ij}} \left(1 - \Pr(o_{ij} = 1) \right)^{1 - O_{ij}}$$

$$\Pr(o_{ij} = 1) = \Pr(P_{ij} = 1)(1 - fn) + (1 - \Pr(P_{ij} = 1))fp$$

New Features in FlyDPI

The screenshot shows the FlyDPI website interface. At the top, there is a logo for the NHRI (New York University) and the text "Drosophila melanogaster Database of Protein Interactomes". There are also icons for "Contact" and "Help". Below the header, there is a navigation bar with "General Search" and "Ping-Pong" tabs. A blue arrow points to the "Ping-Pong" tab. The main search area includes a "Choose high confidence or all" section with radio buttons for "high confidence" and "all". Below this is a "Search the Field" dropdown menu set to "full_text" and a text input field containing "ie. DNA repair, CG10000, CG10001, CT28183". A blue arrow points to the "full_text" dropdown. The "Enter the Keyword" section has a text input field and a blue arrow pointing to it. Below the input field is the text "Multiple items allowed by using comma, ',' as a separator". The "Order" section has a dropdown menu set to "orf". The "select chromosome" section has a dropdown menu set to "ALL". The "select GO category" section has a dropdown menu set to "ALL". The "select expression(multiple)" section has a dropdown menu set to "ALL" and a list of options: "ADULT testes" and "brain". A blue arrow points to this section. The "Select Output Field" section has a grid of checkboxes for various fields: "orf" (checked), "chromosome", "molecular weight", "category", "note", "Unselect all", "expression", "gene_index", "name", "GO_id", "path_name", "locus", "annotation", "lpr_parent_id", "description", "path_title", "transcript", "flybase_note", "Interpro_ID", "ec", and "metabolism". At the bottom, there are "Search" and "Reset" buttons.

Ping-Pong Search

Full-text Search

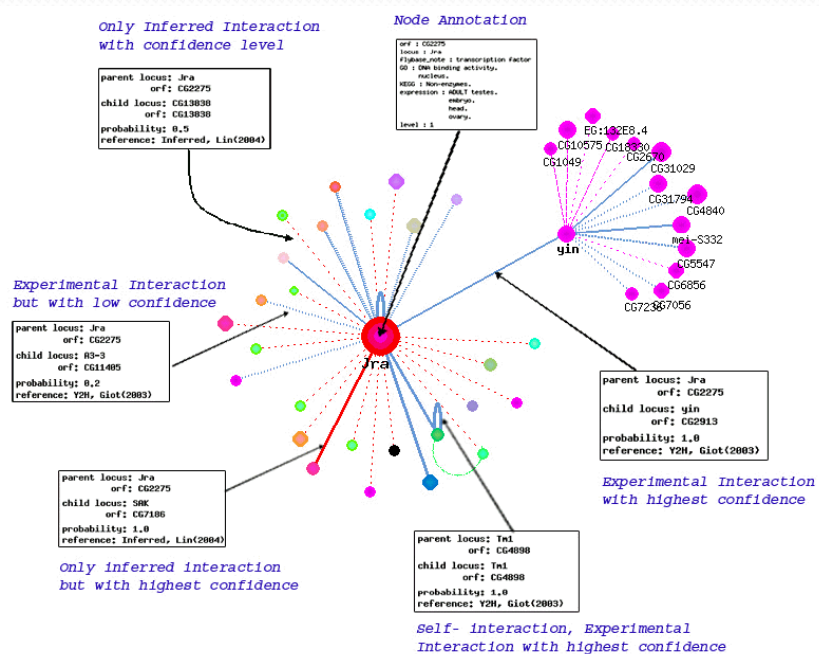
Chromosome Location

Gene Categories from GO

Spatiotemporal Scenarios

Search Results of FlyDPI

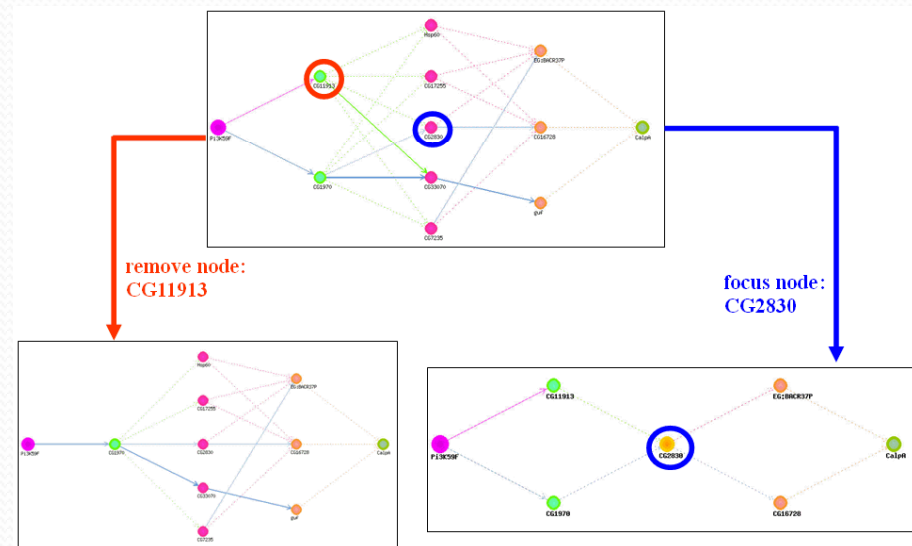
General Search



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

Ping-Pong Search

--Relationship between any two proteins



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

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 Cho¹ ✉, Chia-Ling Chen¹ ✉, Fan-Kai Lin¹ ✉, Chieh-Hua Lin¹ ✉, Pao-Yang Chen¹ ✉, Chen-Zen Lo¹ ✉ and Chao A Hsiung¹ ✉

¹Division of Biostatistics and Bioinformatics, National Health Research Institutes, No. 35 Keyan Rd. Zhunan, Miaoli County 350, Taiwan

²Institute of Information Science, Academia Sinica, No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan

³Institute of Fishery Science, National Taiwan University, No. 1, Sec 4, Roosevelt Road, Taipei, 10617, Taiwan

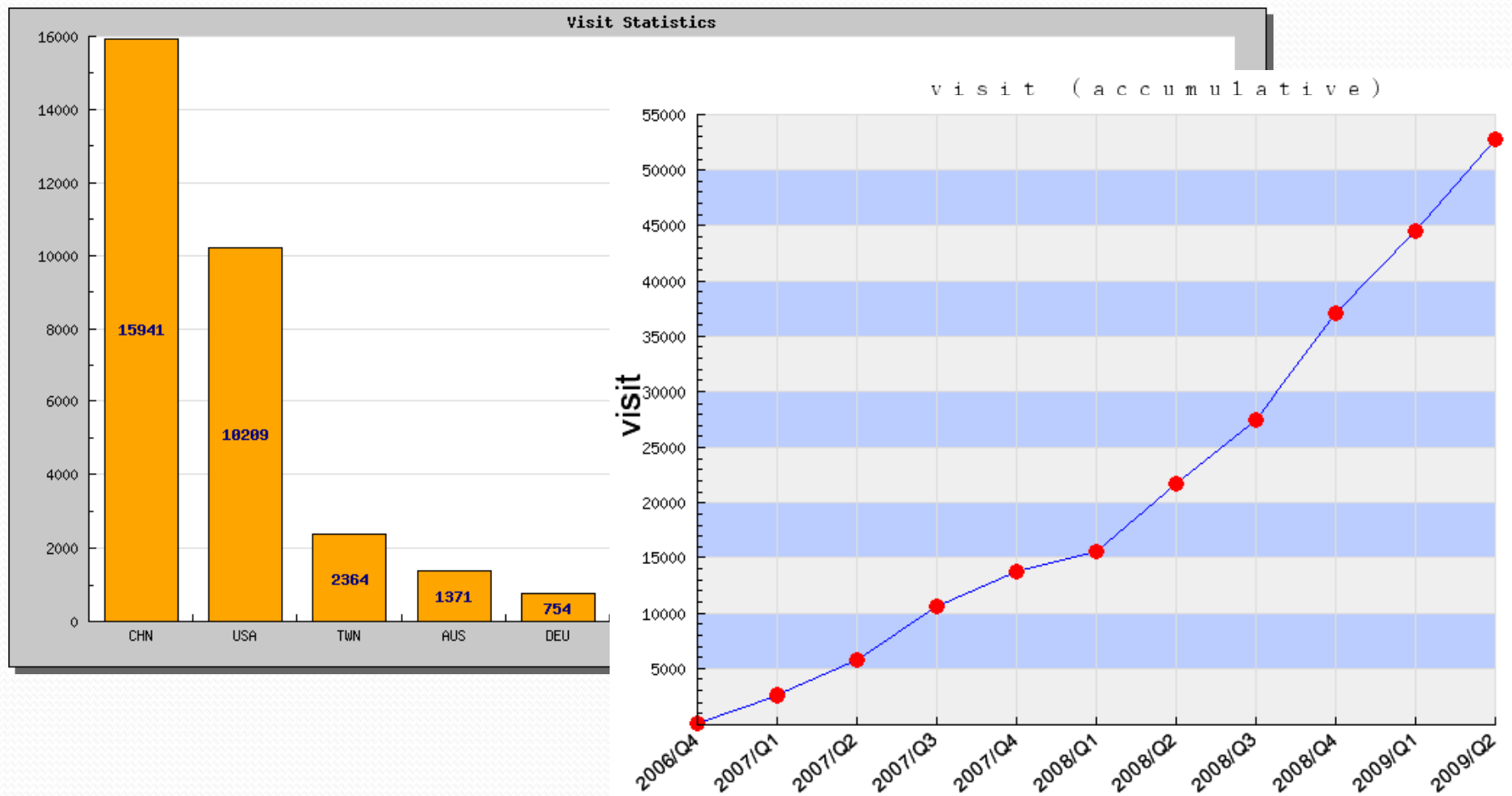
⁴Stem Cell/Regenerative Medicine Program, Genomics Research Center, Academia Sinica., No. 128 Yan-Chiu-Yuan Rd., Sec. 2, Taipei 115, Taiwan

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

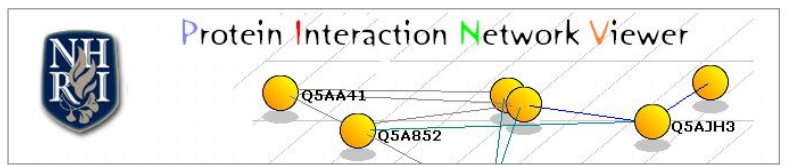
Published 18 December 2006



Visits of FlyDPI (Dec 2006- July 2009)



Next Generation platform for PPI



General Search Ping-Pong Search Submit Network Help & Contact

Uniprot ID

Keyword

Species Drosophila

Probability 0.2

Output Field

<input type="checkbox"/> Select All	<input type="checkbox"/> Unselect All		
<input type="checkbox"/> Accession	<input type="checkbox"/> InterPro ID	<input type="checkbox"/> GO term	<input type="checkbox"/> PFAM ID
<input type="checkbox"/> Protein name	<input type="checkbox"/> InterPro name	<input type="checkbox"/> Ontology	<input type="checkbox"/> PFAM name
<input type="checkbox"/> Gene name	<input type="checkbox"/> InterPro type	<input type="checkbox"/> GO description	<input type="checkbox"/> PFAM clan
<input type="checkbox"/> ORF name			
<input type="checkbox"/> Organism			

© 2005 Division of Biostatistics and Bioinformatics
 National Health Research Institutes, Taiwan. All rights reserved.
 Last update : 2008-05-11

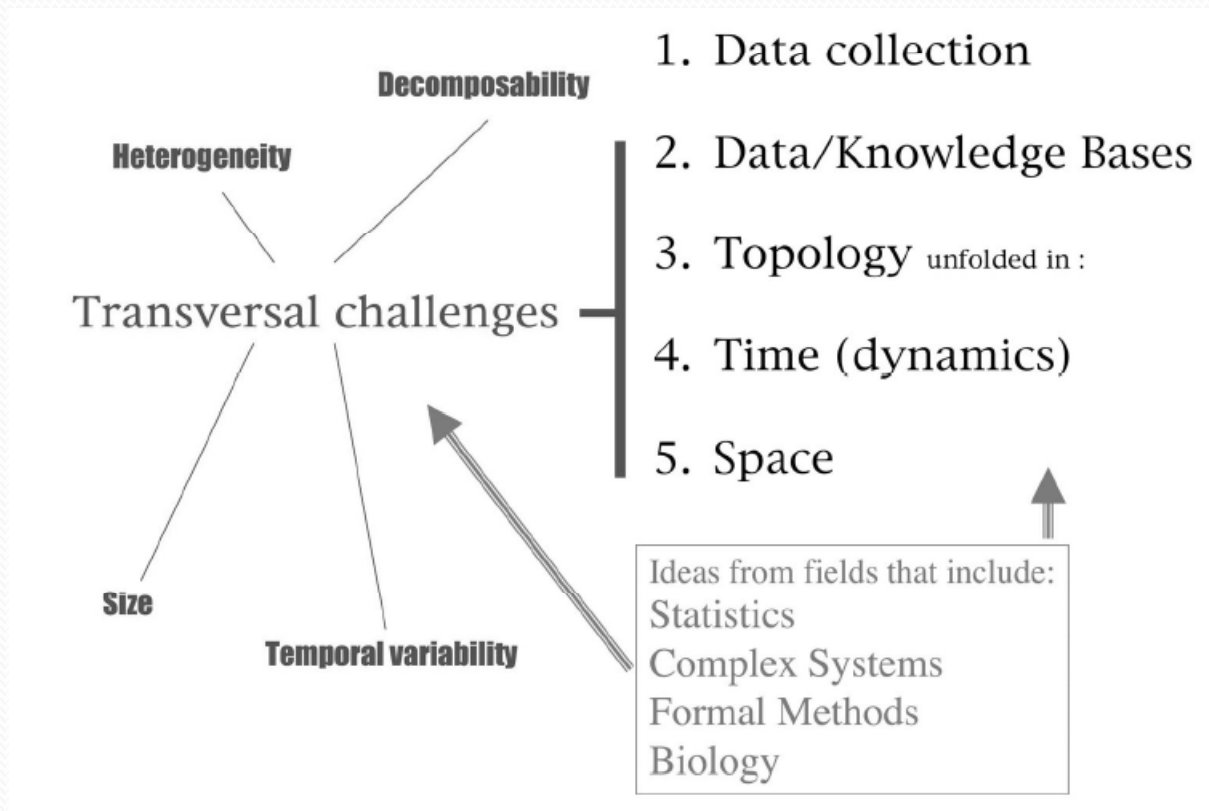
Function Filter Paint Output About

Description:	Structure	GeneOntology	Other
UniProtKB Entry:	Q5A445		
PROTEIN NAME:	Likely ferric reductase		
GENE NAME Name:	CFL2		
ORF names:	CaO19_1264; CaO19_8849		
SOURCE ORGANISM:	Candida albicans		
TAXONOMY ID:	5476 [NCBI, NEWT]		
PROTEIN EXISTENCE:	Predicted		
GENEID	3641811, 3641872		
GO	GO:0016021 , C:integral to membrane GO:0050660 , F:FAD binding GO:0005506 , F:iron ion binding GO:0016491 , F:oxidoreductase activit GO:0006118 , P:electron transport QuickGO		
INTERPRO	IPR013112 , FAD_bd_8, IPR013130 , Fe3_reduct_TM_N, IPR013121 , Fe_red_NAD_bd_6.		
KEGG	cal:CaO19_1264 , cal:CaO19_8849 .		
PFAM	PF08022 , FAD_binding_8,1, PF01794 , Ferric_reduct,1, PF08030 , NAD_binding_6,1.		

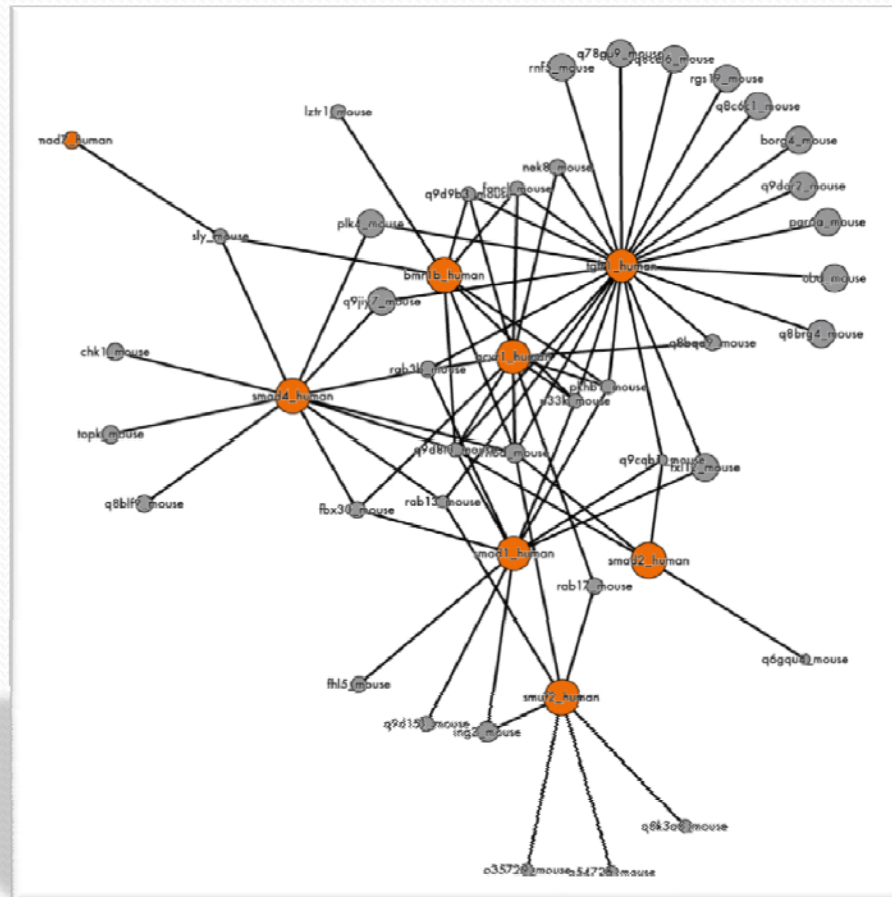
Figure explanation:

Selected Protein	Interact with itself
Centered Protein	Predicted Interact
Interacted Protein	

Challenges Of Biological Networks

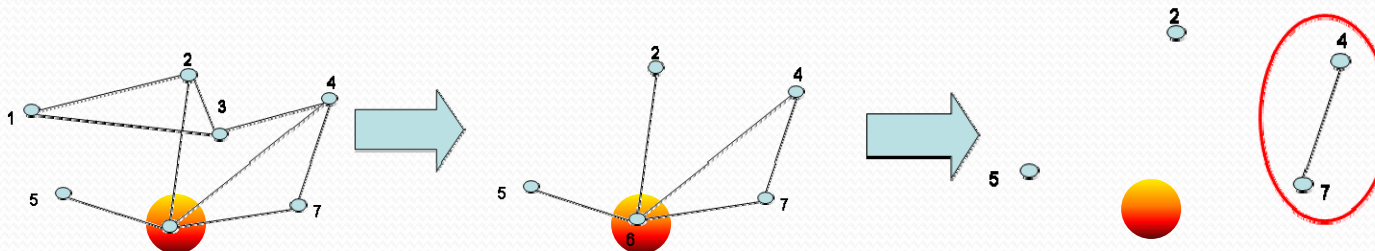
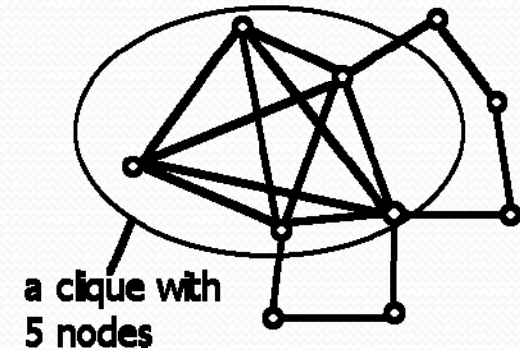


Network Biology: Hub/ Essential Proteins Identification



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



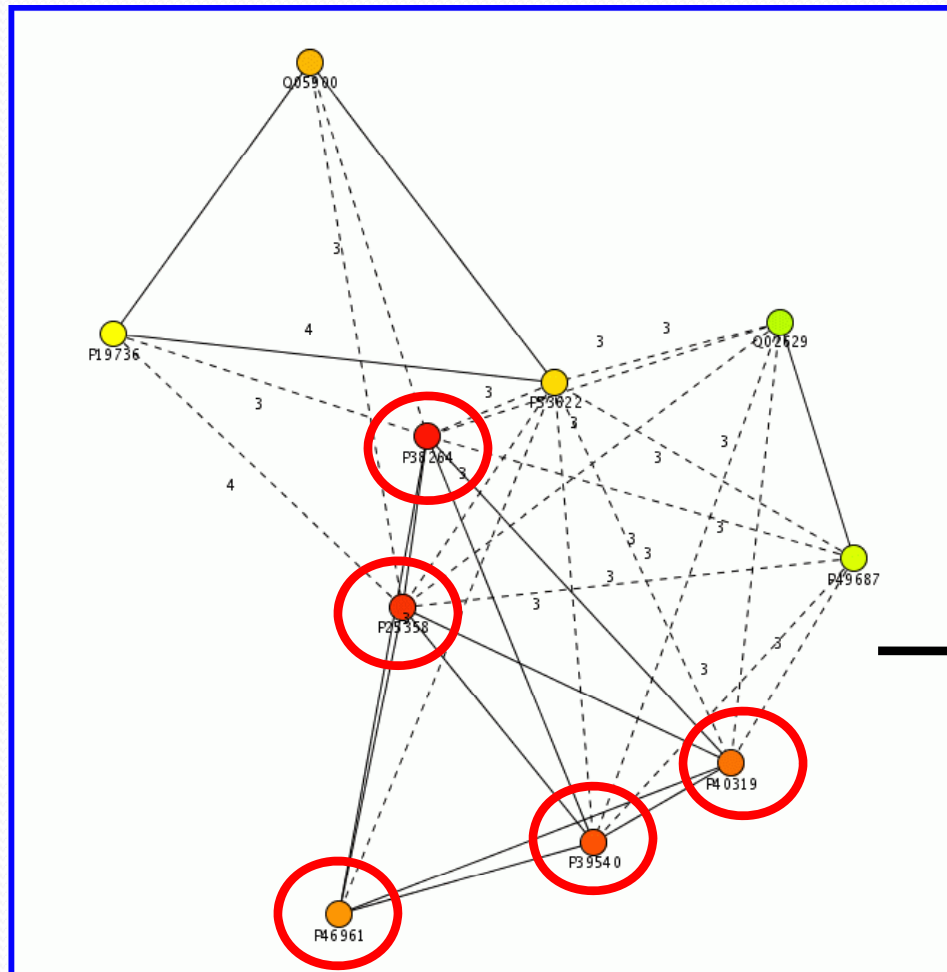
Please input your data and other related information.

Job ID	<input type="text" value="my_job"/> (string with character 0~9, a~z, A~Z)
Input format	<input type="radio"/> PSI <input checked="" type="radio"/> Tab <input type="radio"/> Tab with weight value
Data input	<pre>A56066 S60255 P46414 P10479 P46414 P43063 A41551 A46065 P04278 VENS0F P04925 P10417 P17095 P17947</pre>
Or load it from disk	<input type="text"/> <input type="button" value="Browse"/>
Job note	<input type="text"/>
Email	<input type="text"/> (please use ',' to separate multiple email addresses)

<http://hub.iis.sinica.edu.tw>
NAR 2008 Web issue

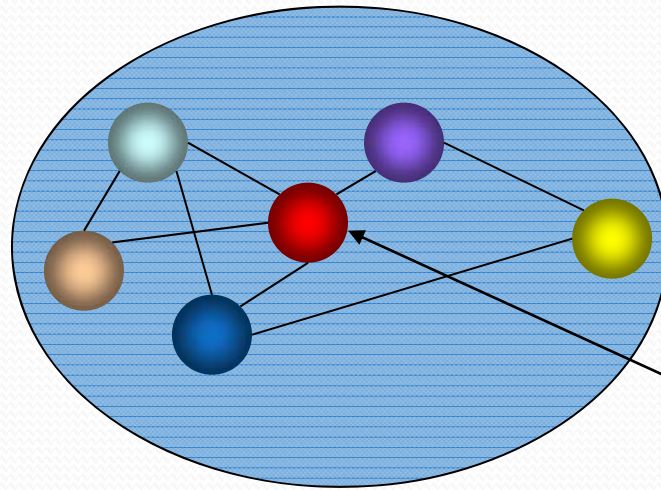


The Relationship of Top 10 in Yeast Complex Network (PPI from DIP, 2007 Jan)



→ Fragile motif in whole network

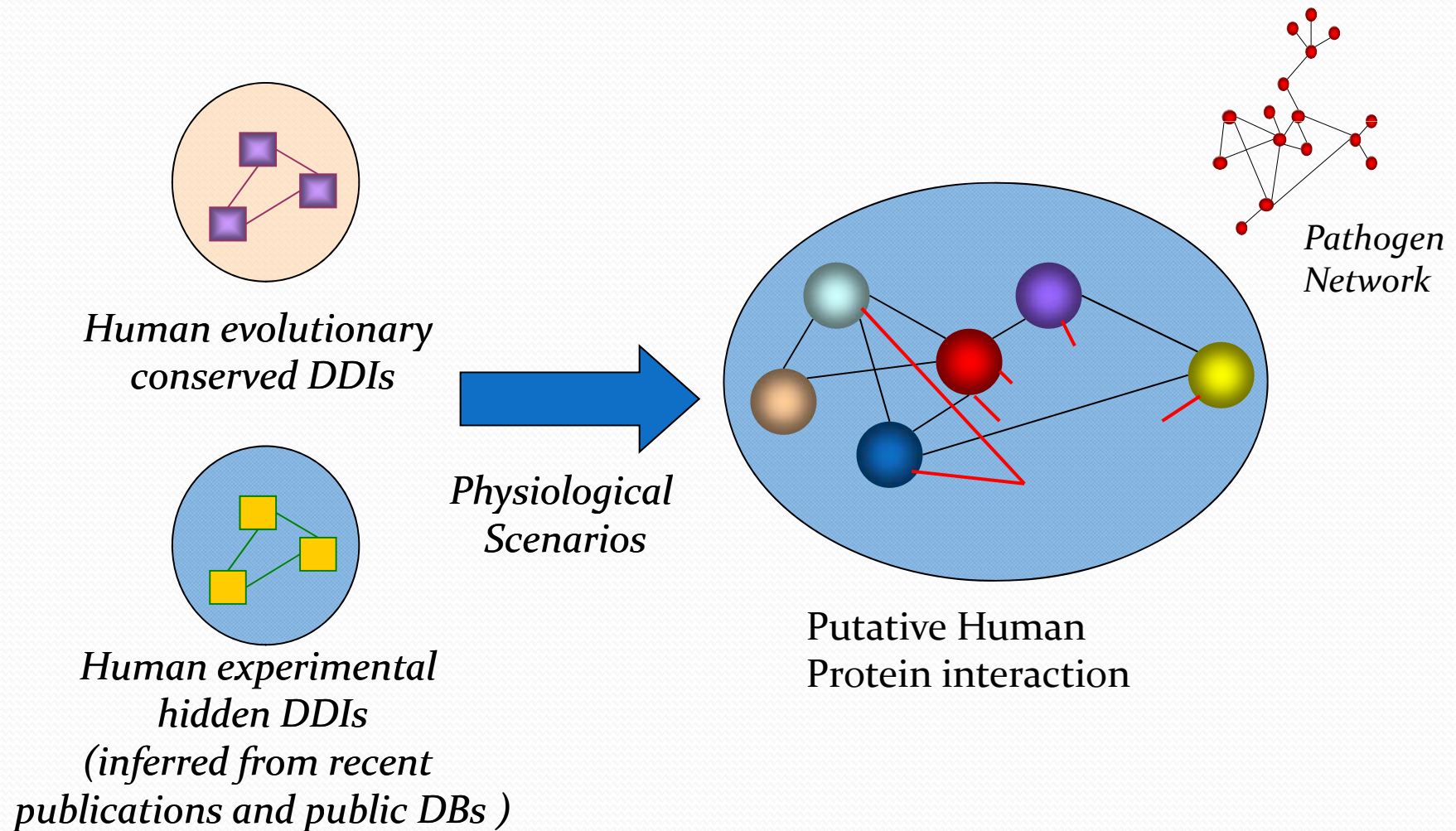
Identification Target Proteins and Hubs for Novel Cancer Therapies



Putative Protein Network in Human Cancer

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

Inferred Protein Interactions by Conserved and hidden DDIs



Human Protein Network

Methodology article

Highly accessed

Open Access

Reconstruction of human protein interolog network using evolutionary conserved network

Tao-Wei Huang¹ ✉, Chung-Yen Lin^{2,3,4} ✉ and Cheng-Yan Kao^{1,5} ✉

¹Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan

²Institute of Information Science, Academia Sinica, Taipei 115, Taiwan

³Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taipei 115, Taiwan

⁴Institute of Fishery Science, National Taiwan University, Taipei 106, Taiwan

⁵Institute for Information Industry, Taipei 106, Taiwan

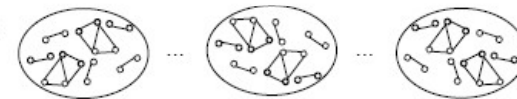
✉ author email ✉ corresponding author email

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



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

PPIs in reference organisms, e.g. mouse, yeast, etc.



PPIs with feature scores (I, D, T, L, P) in human



Interolog using IP and C scores

Confidence score (CS)

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

- Domain-domain combination score (D)

$$D = \sum_{j=1}^{2^m - 12^m - 1} \sum_{i=1}^{N(pd_i, pd_j)} \frac{N(pd_i, pd_j)}{N(pd_i, pd_j)} \text{ if } pd_i \in PD_d, pd_j \in PD_d$$

- Tissue specificity score (T)

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

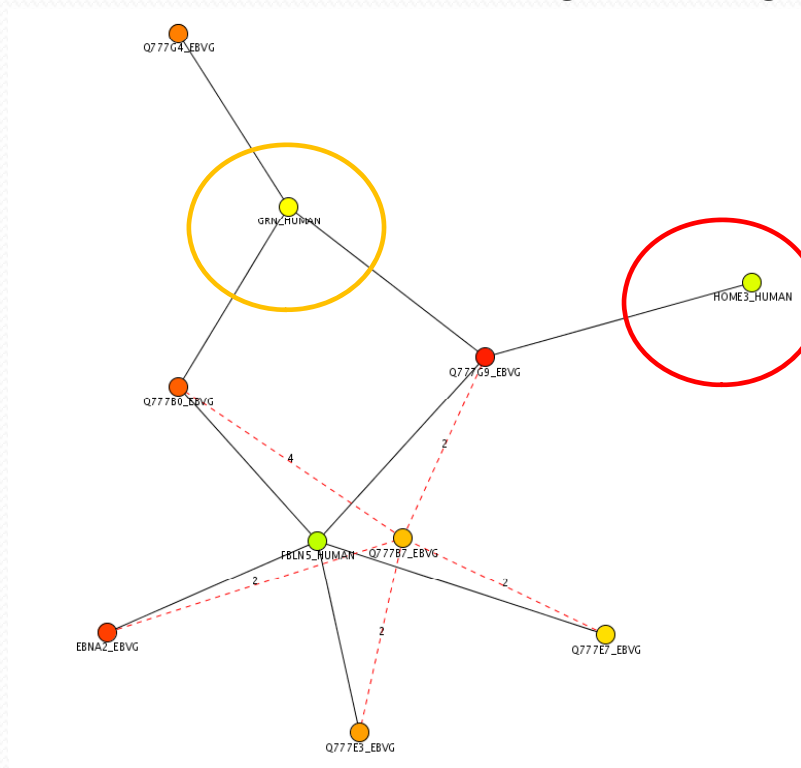
- Sub-cellular localization score (L)

- Cell-cycle stage score (P)

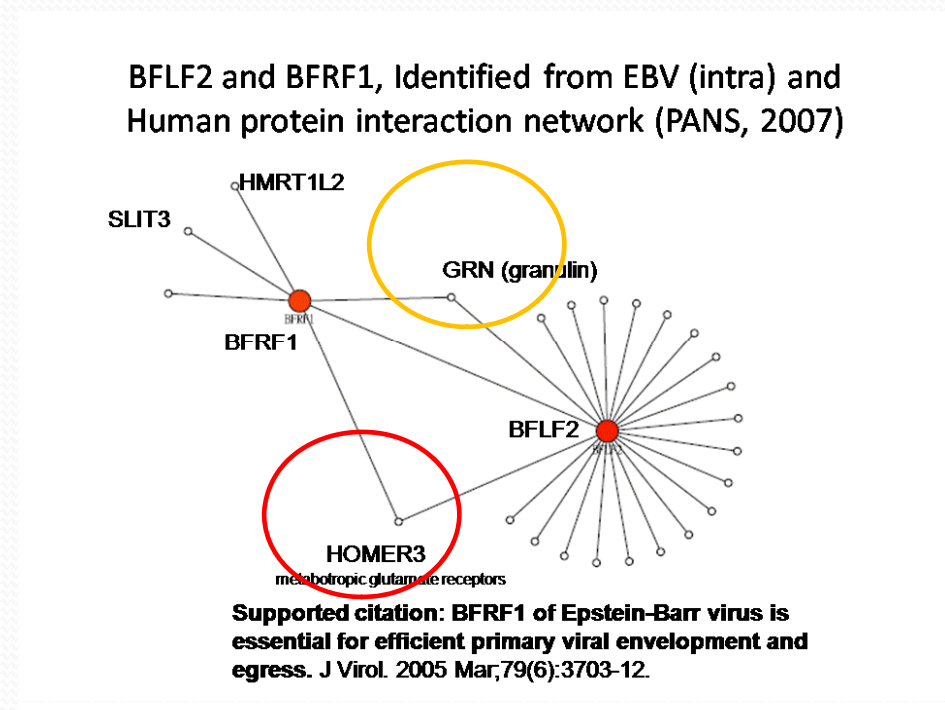
$$CS = w_I * \frac{I}{I_R} + w_D * \frac{D}{D_K} + w_T * \frac{T}{T_K} + w_L * \frac{L}{L_K} + w_P * \frac{P}{P_K}$$

Interactome among Pathogens and Host

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



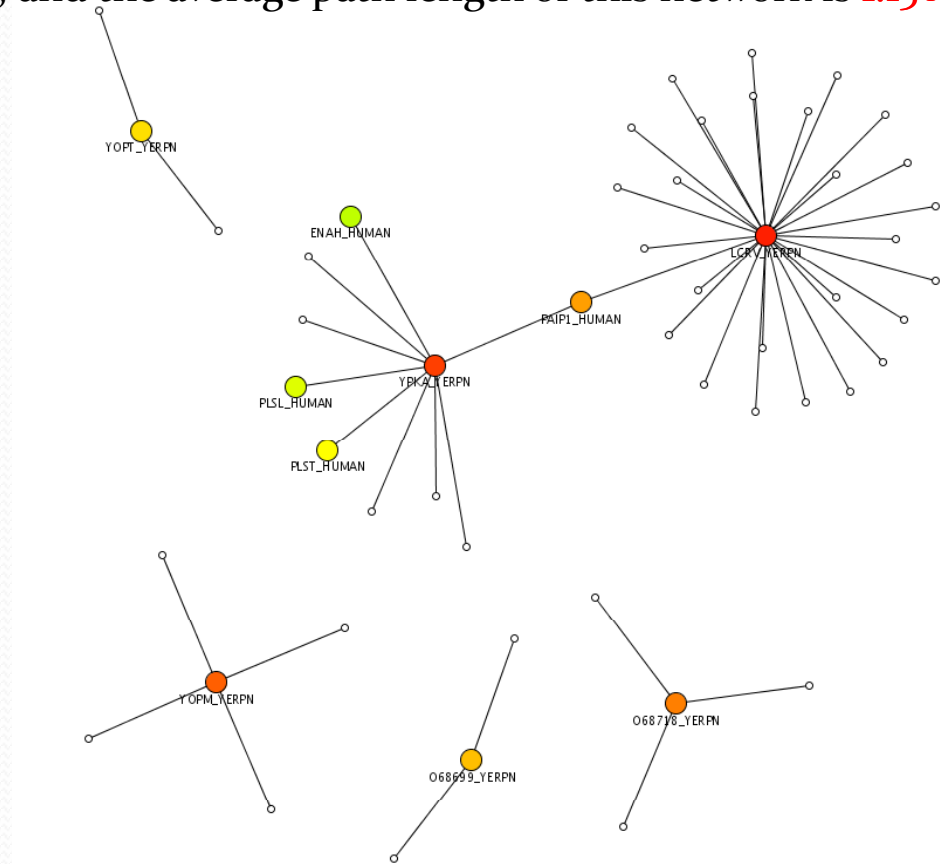
Source: EBV and Human, Dyer *et al.*, 2008



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

Interactome of *Yersinia pestis* and Human host

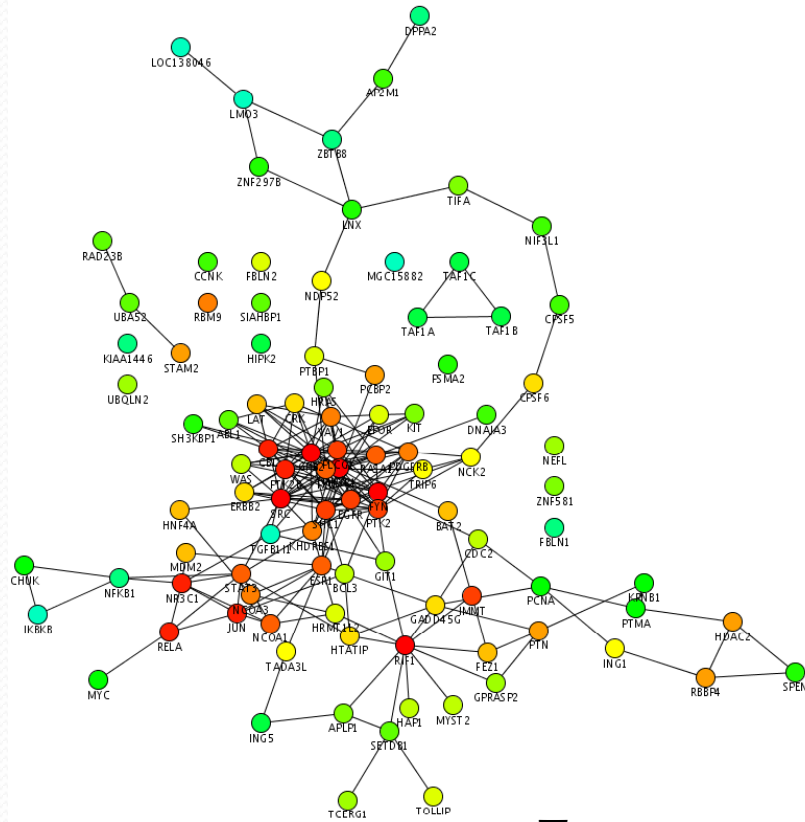
There are **56** nodes and **49** edges in your network. The clustering coefficient of this network is **0**, and the average path length of this network is **1.15649**.



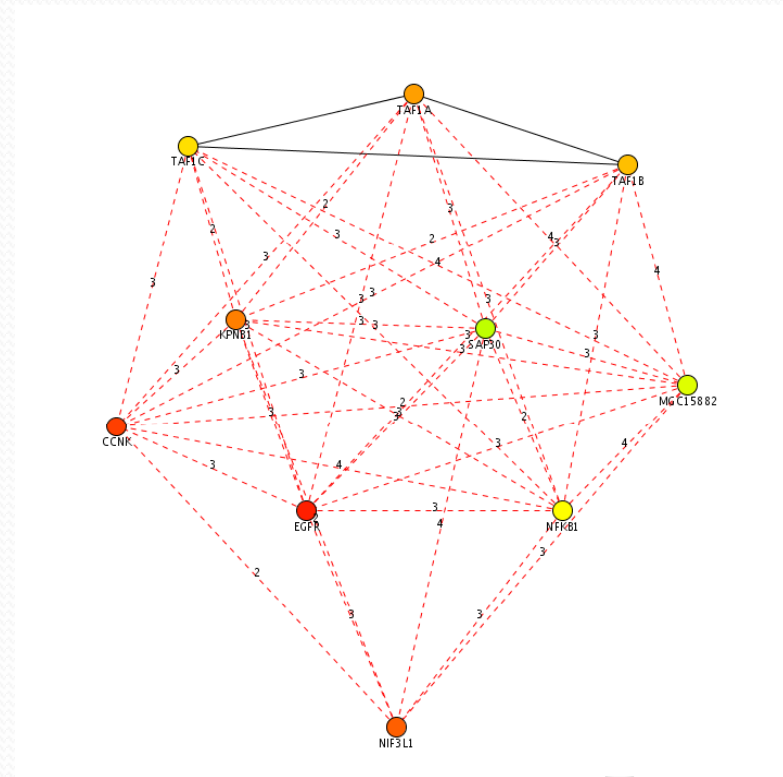
http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_30_02_36_24

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

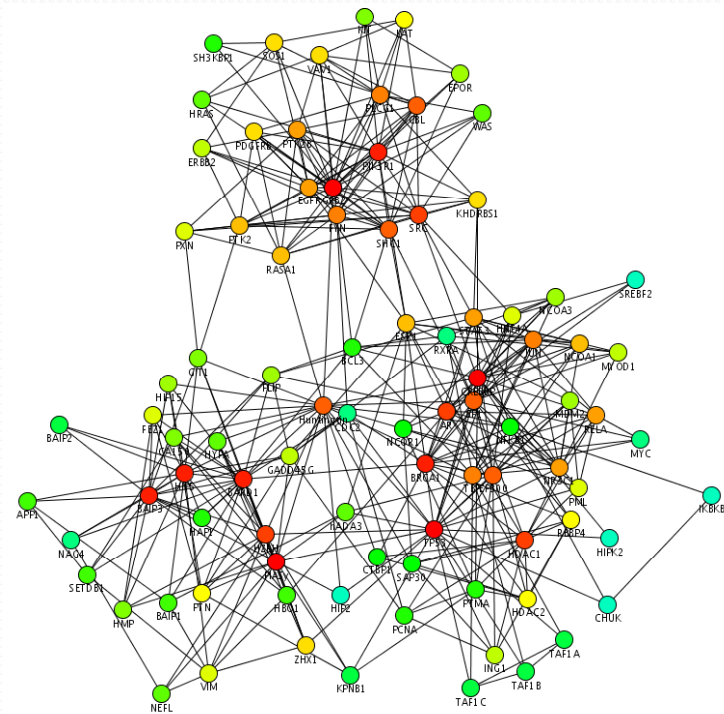


Top 10

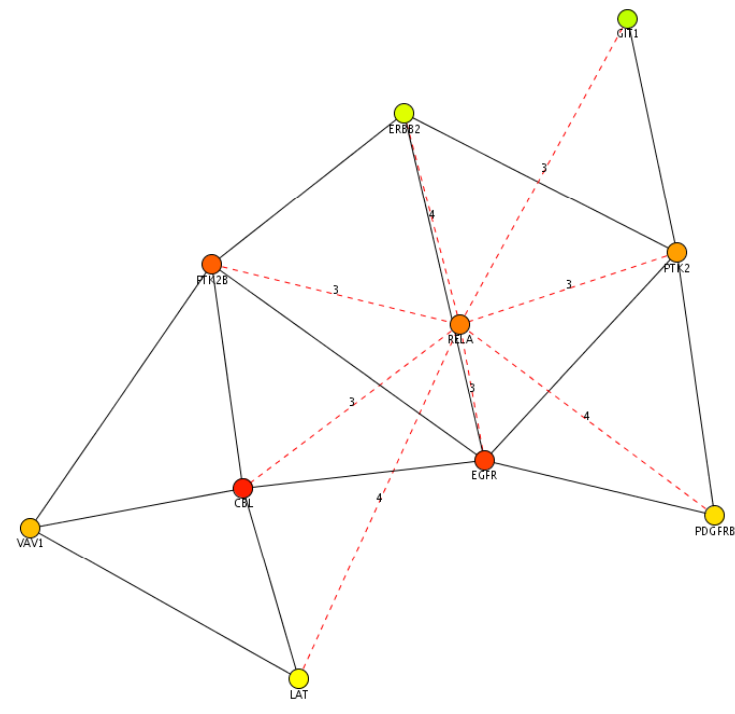
http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_20_00_44_35

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 coefficient of this network is **0.23954**, and the average path length of this network is **2.85459**.



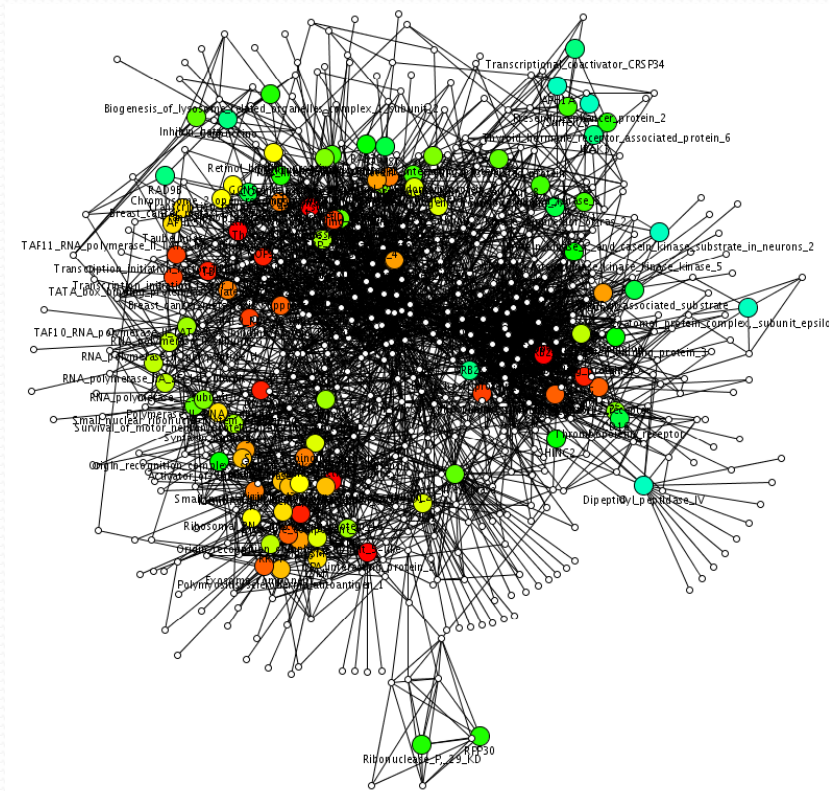
Top 90



Top 10

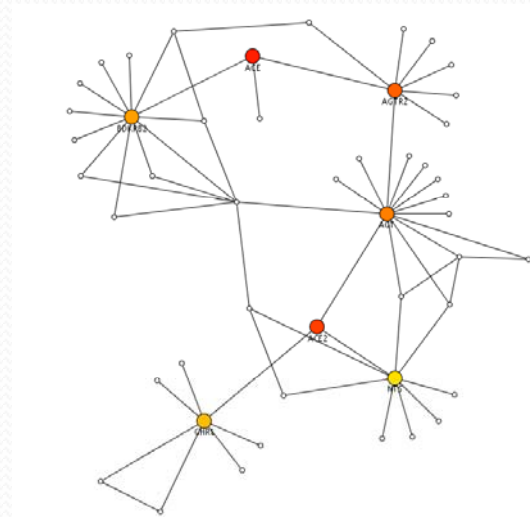
http://hub.iis.sinica.edu.tw/Hubba/result.php?ID=upload/2008_04_16_23_12_26

Extract Sub-network with Targeted Genes



Targeted Genes

ACE
ACE2
AGTR2
AGT
BDKRB2
GHRL
NTS
AGTR1
ARRB2



Relationships among these proteins

Human Interactome, HPRD, 2007/9/01

Hubba plug-in for Cytoscape

Control Panel

Network: galFiltered.sif (331 nodes and 362 edges)

Select nodes: Hubba nodes

Choose ranking method: DSS

Top group: 10

Particular nodes: Choose particular group

Display options: Check the first-stage nodes, Display the shortest path

Submit

galFiltered.sif_DSS_top10_with_neighbors

Results Panel

Network: galFiltered.sif

Ranking method: DSS

Top 10

Rank	Node
1	YPL248C
2	YMR04...
3	YCL067C
4	YOL051W
5	YDR244W
6	YBR020W
7	YGL035C
8	YHR084W
9	YGL153W
10	YPL1000W

Save Current Rank

Data Panel

ID	Pathway	annotation.GO BIOLOGICAL_PROCESS	annotation.GO CELLULAR_COMPONENT
YMR043W	[KEGG pathway: MAPK signaling pathway]	[DNA replication initiation, cell cycle, regulation of tr...	[nuclear chromatin, nucleus]
YJL194W	[KEGG pathway: Cell cycle - yeast]	[DNA replication, G1/S transition of mitotic cell cycle...	[DNA replication preinitiation complex, membra
YFL026W	[KEGG pathway: MAPK signaling pathway]	[G-protein coupled receptor protein signaling pathw...	[integral to membrane, integral to plasma mem
YBR160W	[KEGG pathway: Cell cycle - yeast]	[G1/S transition of mitotic cell cycle, G2/M transition ...	[actral microtubule, cellular bud neck, cyclin do
YAL040C	[KEGG pathway: Cell cycle - yeast]	[G1/S transition of mitotic cell cycle, cell cycle, cell d...	[nucleus]
YER111C	[KEGG pathway: Cell cycle - yeast, KEGG pathway: ...	[G1/S transition of mitotic cell cycle, regulation of tra...	[nucleus]
YGR108W	[KEGG pathway: Cell cycle - yeast]	[G2/M transition of mitotic cell cycle, cell cycle, cell d...	[cytoplasm, nucleus]
YPR110W	[KEGG pathway: Cell cycle - yeast]	[G2/M transition of mitotic cell cycle, cell cycle, cell d...	[cellular bud neck, cytoplasm, nucleus, spindle

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser

Transcriptomics: Primers and microarray probe design tools in high throughput Experiments

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



Genomics Studies For High Throughput Research : PDA

PDA
Primer Design Assistant

Help Contact

Input format: fasta text

Sequence(s) input or file upload

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%
T _m	T _m of forward and reversed primers restricted to be higher than 50°C
ΔT _m	restricted to be smaller than 5°C

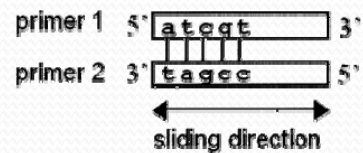
- Advanced options

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.

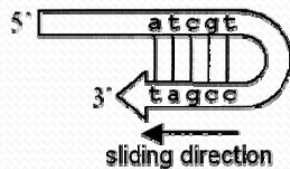
$$T_m(^{\circ}\text{C}) = 59.9 + (0.41 \times \text{GC content}) - \left(\frac{675}{\text{primer length}} \right)$$

Calculation of the Stability of DNA Duplexes

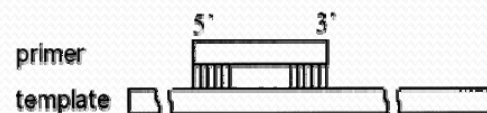
A Primer-to-primer annealing



B Hairpin structure



C Primer-to-template annealing



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	-0.92
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



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

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):

$$R = 100 - \Delta(T_m) + \Delta G_{\text{forward}}^{\circ}(3' - 5') + \Delta G_{\text{reverse}}^{\circ}(3' - 5') + \text{hairpin score} + \text{dimer score}$$

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

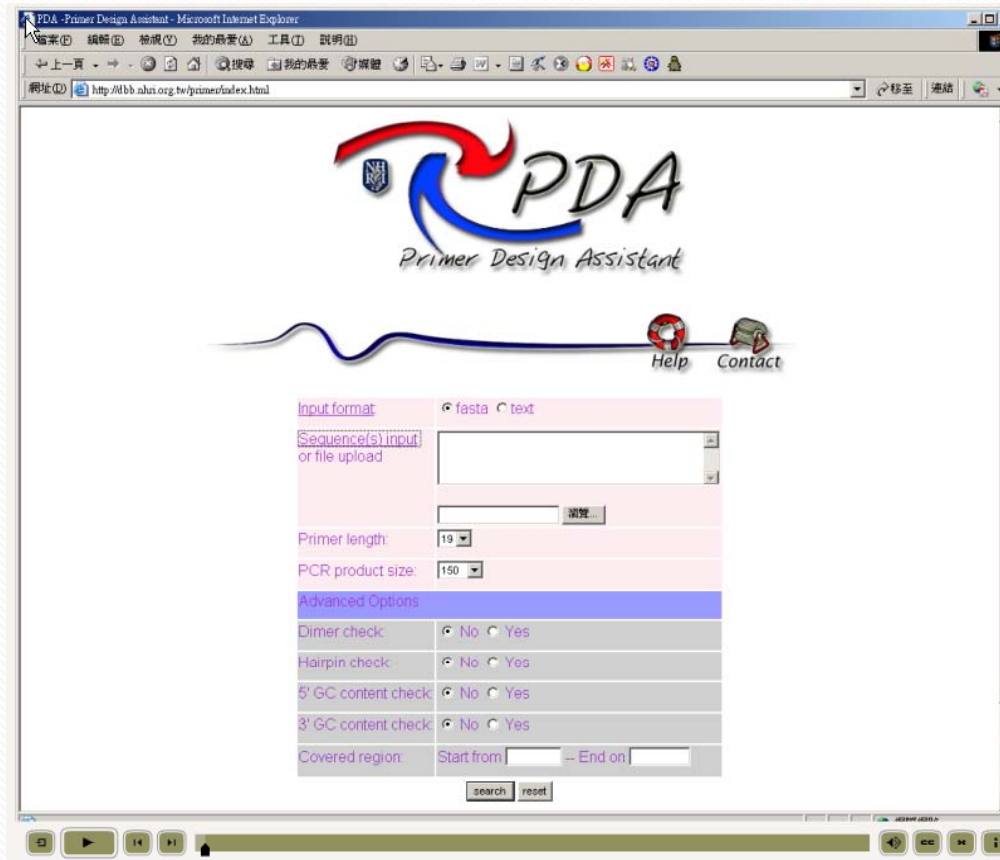
Currently available service (conti)



Input format:	<input checked="" type="radio"/> fasta <input type="radio"/> text
Sequence(s) input or file upload	<input type="text"/> <input type="text"/> 瀏覽...
Primer length:	19 ▼
PCR product size:	150 ▼

Batch primer
design for
unified
experimental
conditions

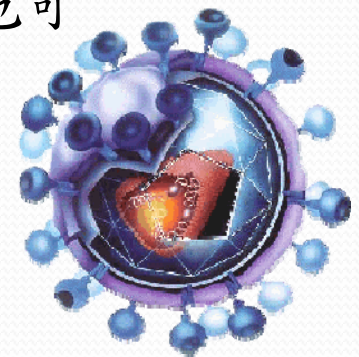
Flash Demo for PDA



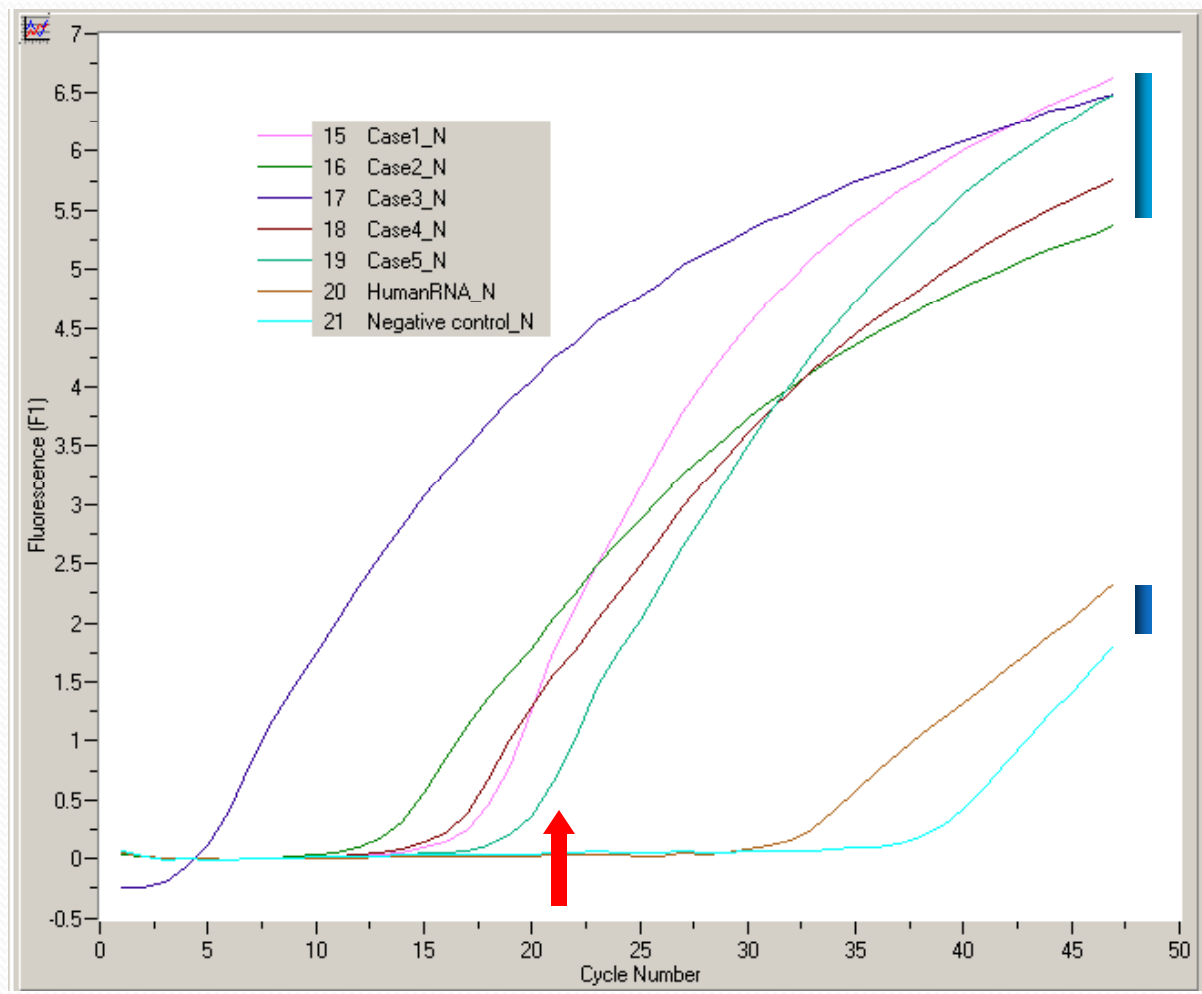
http://dbb.nhri.org.tw/primer/20080726_PDA_music.htm

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

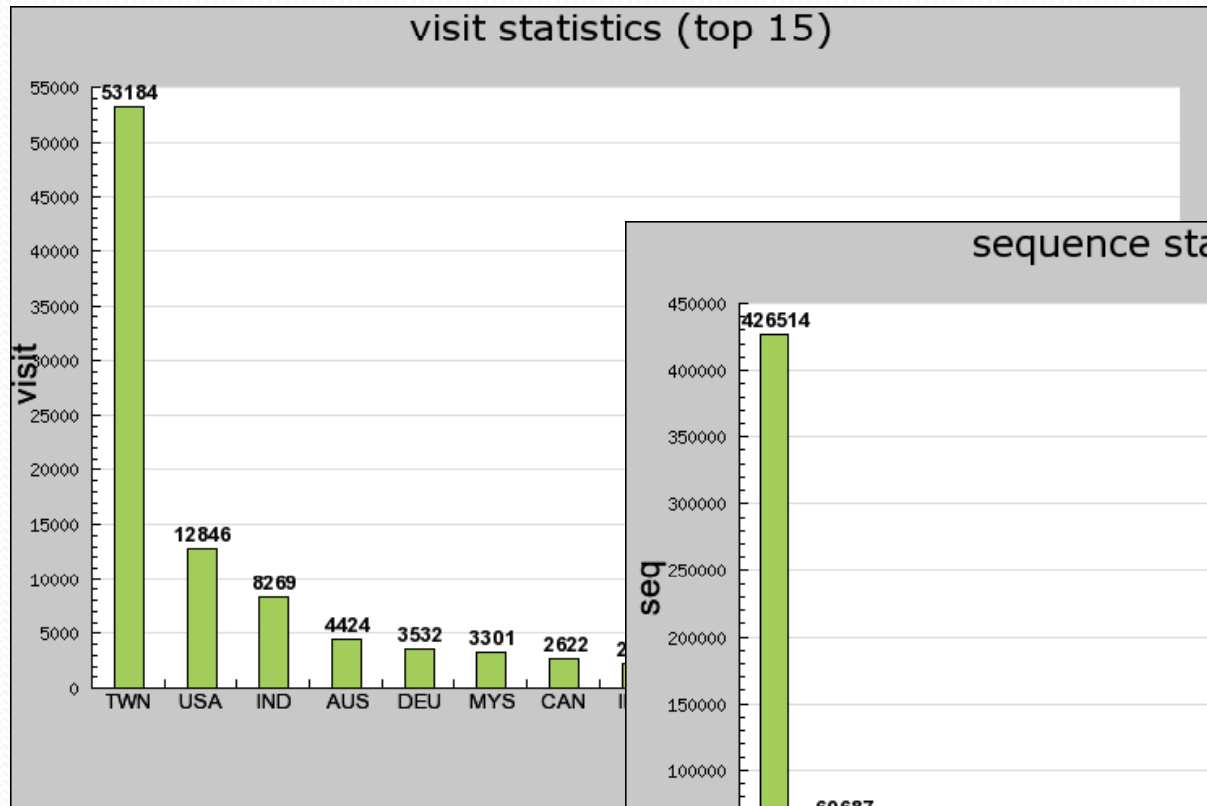
(-) Control



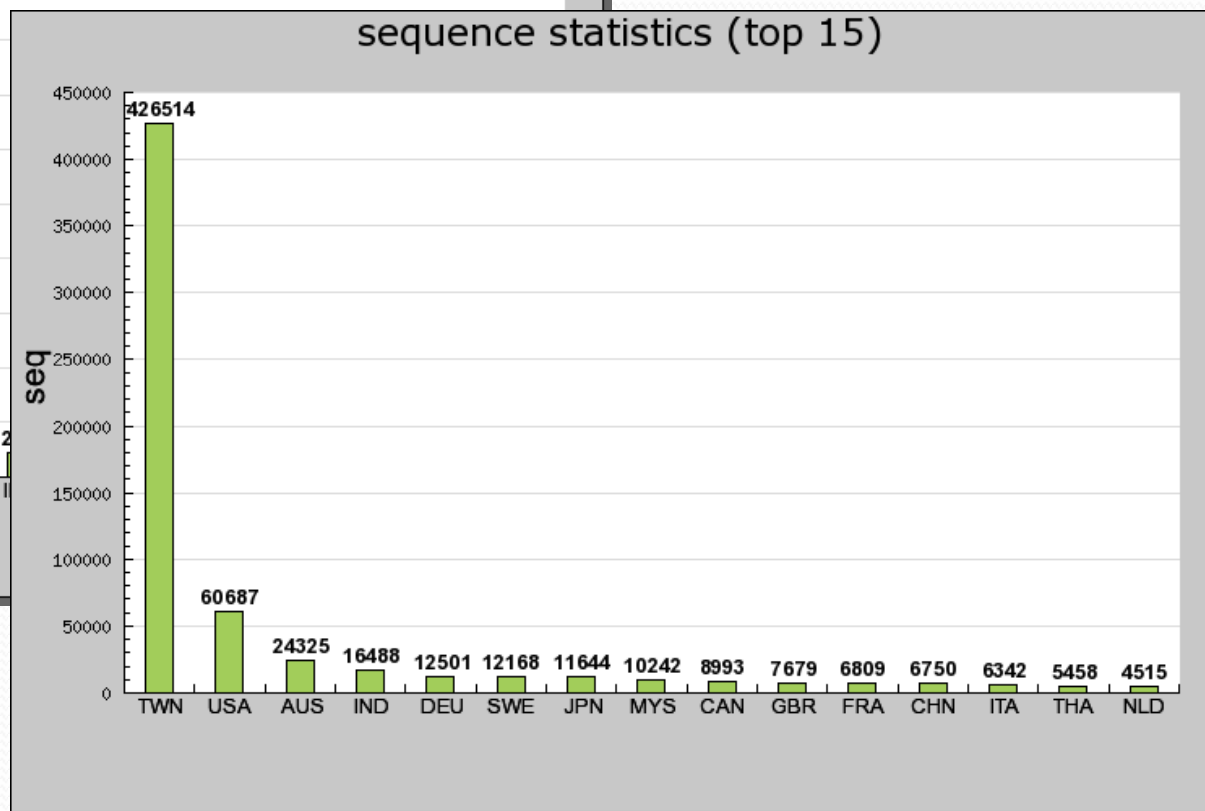
Primer Design Assistant (PDA)



visit statistics (top 15)



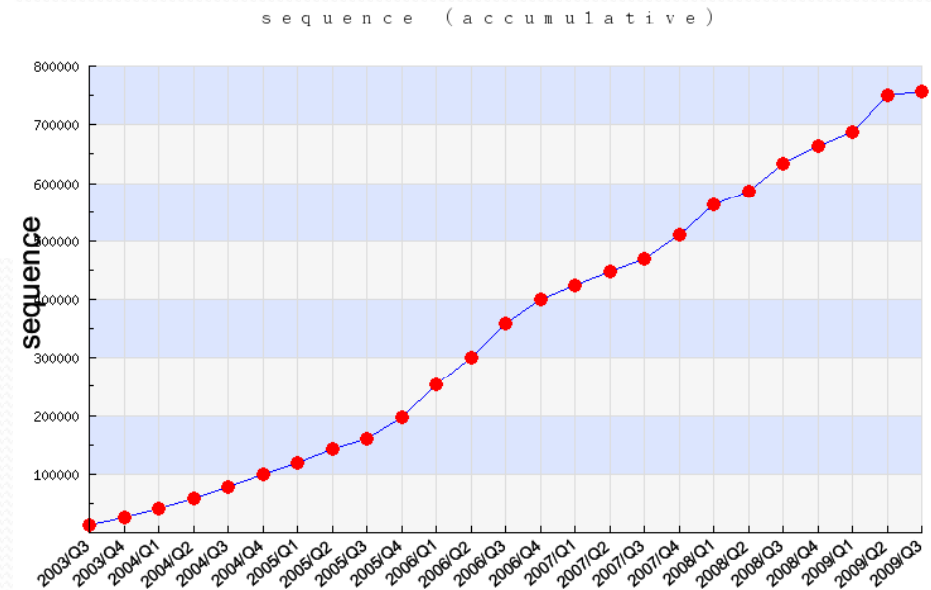
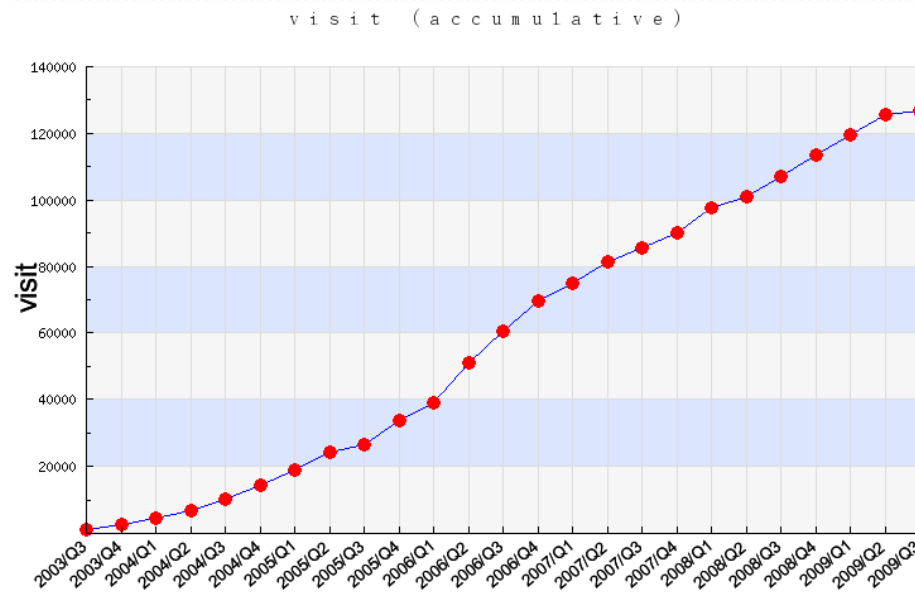
sequence statistics (top 15)



From July 2003 to May 2009

Nucleic Acids Res. 2003, 31: 3751-3754

Primer Design Assistant (PDA)



From July 2003 to July 2009
Over 125,000 Visits and 750,000 submitted Seqs

Nucleic Acids Res. 2003, 31: 3751-3754

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

Received: 17 June 2005

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

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/>]

tribution License (<http://creativecommons.org/licenses/by/2.0>), the original

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.

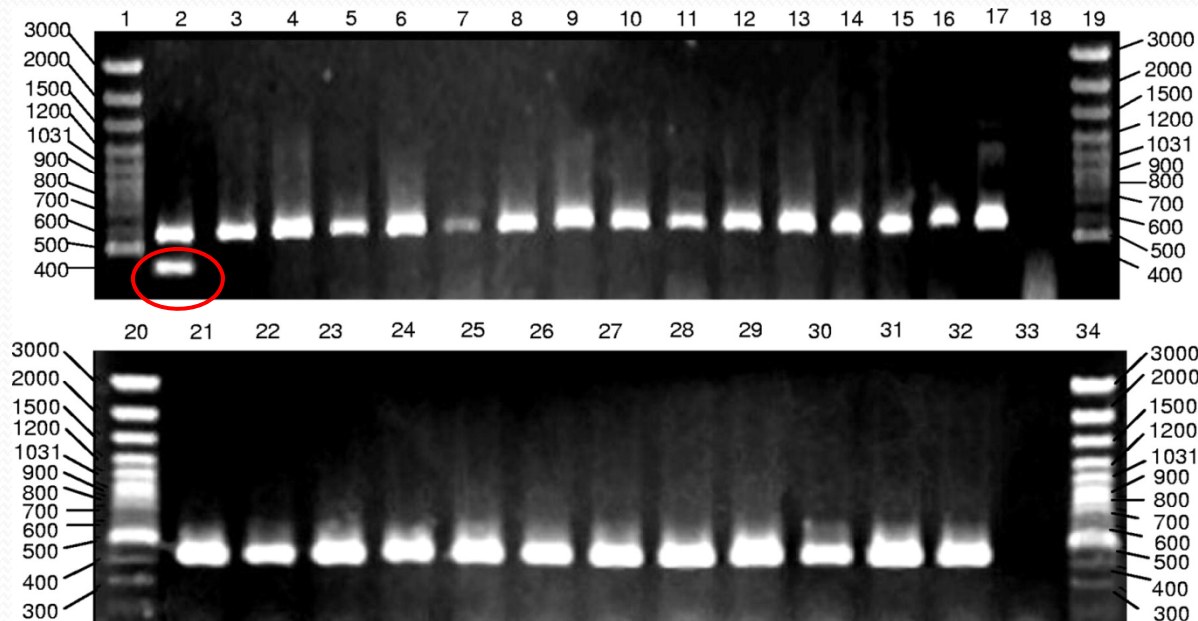


Specific Primers for Rapid Detection of *Microsporium audouinii* by PCR in Clinical Samples[∇]

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, Portugal²

Received 11 April 2006/Returned for modification 3 July 2006/Accepted 12 September 2006

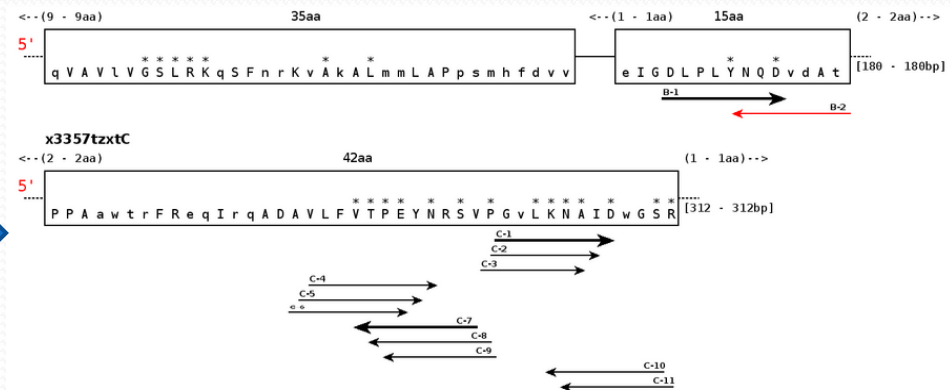
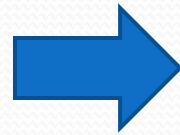


Design Degenerate Primers : iCODEHOP

- PCR primers designed from protein multiple sequence alignments
- The interactive program for creating COnsensus DEgenerate Hybrid Oligonucleotide Primers (iCODEHOP)

Aligned Protein sequences

	1	5	10	15	20	25
<i>α-DTX</i>	E	P	R	R	L	C
<i>DTX1</i>	Q	P	L	R	K	L
<i>DTXK</i>	--	A	A	K	Y	C
<i>β-DTX</i>	--	A	A	K	Y	C
<i>BPH</i>	--	R	P	D	F	C



Genomics Studies For High Throughput Research : UPS



BMC Bioinformatics 2007, 9(Suppl 1):S8

UPS
Unique Probe Selector

Home Demo Help Contact

Probe Uniqueness Unique Probe within group Unique Probe in the specific organism

Sequence (s) Paste or File upload

Probe Length DEMO

Probe # for each sequence (maximum 3)

Job note (optional)

E-mail (optional) (Recommended)

submit reset

©2008 Institute of Information Science, Academia Sinica. All rights reserved.

Latest updated: 2008/12/08

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

Basic Criteria for Probe Selection

- ① Probe length: from 30 ~ 120 bps
- ② Melting Temperature
 - ✓ The probe annealing temperature (T_a) is determined by melting temperature (T_m). Probe T_m depends on several physiochemical factors and is calculated in the following equation based on Nearest Neighbor model

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

- 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 non-target template
 - ✓ Here we used Li *et al*'s (NAR 2003) experimentally established criteria to exclude unsuitable oligonucleotides: **identity of $\geq 85\%$, 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*	<input checked="" type="radio"/> Unique Probe within group <input checked="" type="radio"/> Unique Probe based on the specific organism <input type="radio"/> Unique Probe based on user's defined organism
	<input type="text" value="Aedes_aegypti (yellow fever mosquito)"/> <input type="button" value="Browse"/>
Sequence (s) Paste or File upload*	<input type="text"/> <input type="button" value="Browse"/> DEMO
Probe Length	70
Probe # for each sequence	1 (maximum 3)
Job note (optional)	<input type="text"/>
E-mail*	<input type="text"/>
Advanced Options	
[Salt]	salt_conc 0.58 (0~1M)
Degenerate probe allowed	<input type="radio"/> yes <input checked="" type="radio"/> no
	<input type="button" value="submit"/> <input type="button" value="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_ramidii
- Helianthus_annuus (sunflower)
- Hydra_magnipapillata
- Homo_sapiens (human)

Jobs Accepted by UPS



Home



Demo



Help



Contact

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>

Add to my favorite

Thanks for using UPS. Any comment will be appreciated.

Your faithfully.

UPS Administrator.

Output of UPS

Job Note : PTP family
Type of Probe Uniqueness : Unique Probe within group

Page 1

Output for UPS





Total : 111

Advanced Options filter

Sequence_ID	Rank	CG%	Tm	probe sequence	delta G
ABL	1	56	73	ctgagcaggttgatgacaggggacacctacacagcccatgctggagccaagtccccatcaaatggactg	0.183
ARG	1	40	68	gagccaaattcctattaagtggacagcaccagagagcttgcctacaataccttctcaattaaatctga	1.154
EGFR	1	44	69	gcagaaggaggcaaagtgcctatcaagtggatggcaltggaatcaattttacacagaatctatacccacc	-1.485
TNK1	1	76	78	tggtcggcctctggcggtgcccggggcgctacgtcatggcgggcccccgcctatcccctacacctg	-3.79
TXK	1	40	68	agccaagttcccaatcaagtggccccctctgaagttttctttcaataagtacagcagtaaatctgat	0.802
TYK2	1	69	77	cctagccaaggccgtgcccgaaggccacgactactaccgctgcgcgaggatggggacagcccgtgttc	-1.649
TYRO3	1	54	73	tcggactctcccgaagatctacagtgggactactatcgtcaaggctgtgcctccaaactgcctgtcaa	-0.65
VEGFR1	1	44	69	gccttcccgggatattataagaaccccgattatgtgagaaaaggagatactcgacttctctgaaatg	0.096
VEGFR2	1	44	70	gcccgggatattataagatccagattatgtcagaaaaggagatgctcgcctccccttgaatggatgg	0.232
VEGFR3	1	66	75	gccttcccgggacatctcaaaagaccccgactactcgcgaaggcagtgcccgggtgccctgaaatg	-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 No: ...
Type: ...

Page 1

Advanced Options Filter

GC% from % to %

Tm range not lower than °C

Total : 111

- Sequence_ID
- ABL
 - ARG
 - EGFR
 - TNK1
 - TXK
 - TYK2
 - TYRO3
 - VEGFR1
 - VEGFR2
 - VEGFR3

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	ctgagcagggtgatgacaggggacacctacacagcccatgctggagccaagtccccatcaaatggactg	0.183
ARG	1	40	68	gagccaaatttcctattaagtggacagcaccagagagtcttgccacaataccttctcaattaaatctga	1.154
EGFR	1	44	69	gcagaaggaggcaaagtgcctatcaagtggatggcattggaatcaattttacagaaatctatacccacc	-1.485
TXK	1	40	68	agccaagtccaatcaagtggccccctcctgaagttttctttcaataagtcagcagtaaatctgat	0.802
TYRO3	1	54	73	tcggactctcccgaagatctacagtgggactactatcgtaaggctgtgcctccaaactgcctgtcaa	-0.65

- We provide
1. Best probe
 2. All probes
 3. All probes
 4. In silico hybridization

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
agttcccacatcaaatggactgcacccgagagcctggcctacaacaagtctctcca
>ARG_1
tcctattaagtggacagcaccagagagtcttgccctacaataccttctcaattaa
>EGFR_1
aagtgcctatcaagtggatggcatggaatcaattttacacagaatctatacc
>TNK1_1
acttcgggctgggta
>TXK_1
atgacaaggtacgt
>TYK2_1
gtactaccgcgtgc
>ABL_1
agttcccacatcaaatggactgcacccgagagcctggcctacaacaagtctctcca
>ABL_78
gttcccacatcaaatggactgcacccgagagcctggcctacaacaagtctctcca
>ARG_1
tcctattaagtggacagcaccagagagtcttgccctacaataccttctcaattaa
>ARG_2
atttcctattaagtggacagcaccagagagtcttgccctacaataccttctcaattaa
>ARG_78
cctattaagtggacagcaccagagagtcttgccctacaataccttctcaattaa
    
```

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	CCinside	TMinside	Probe sequence
TC12005	1	6	53	77	Y	Y	ctgcccgcggtctttcgggttcggcgccga
TC12005	2	5	51	77	Y	Y	ttcgggtctctcggcctaacgagaggttc

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. Schaeffer,
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

Sequences producing significant alignments:

              Score  E
              (bits) Value
>Hs.431048                139  7e-033
Length = 5881

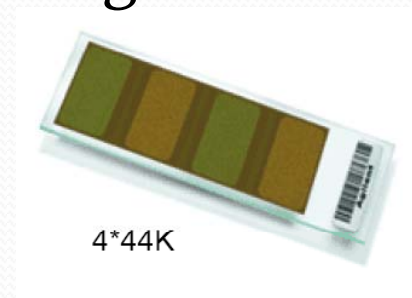
Score = 139 bits (70), Expect = 7e-033
Identities = 70/70 (100%)
Strand = Plus / Plus

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

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

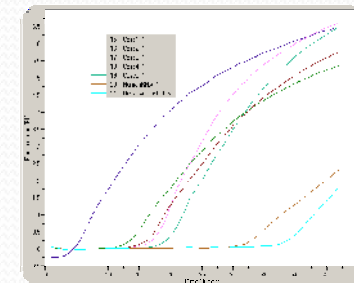
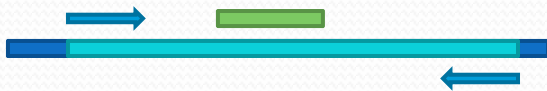
Applications of UPS

- Probe design for hybridization in low and high throughput experiments
 - Shrimp, WSSV EST array (7K*6)
 - Human whole genome array (44K)
 - Array design for transcriptome from high throughput sequencer
 - Allow user to upload their own genetic background for probe design to compensate those genome not listed in Unigene



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



Phylogenetic Analysis:

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

林仲彥 蘇聖堯



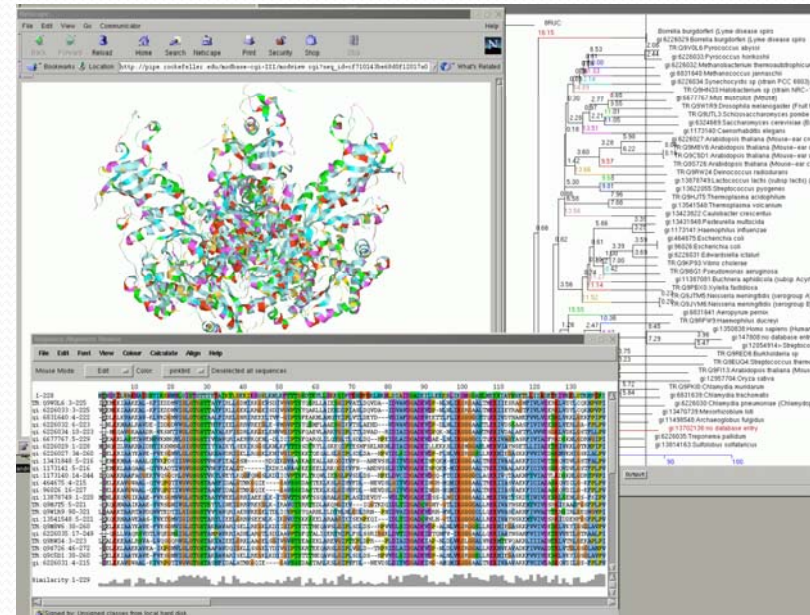
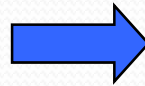
中央研究院資訊科學研究所

May 15, 2009

Coding Characters and Defining Homology



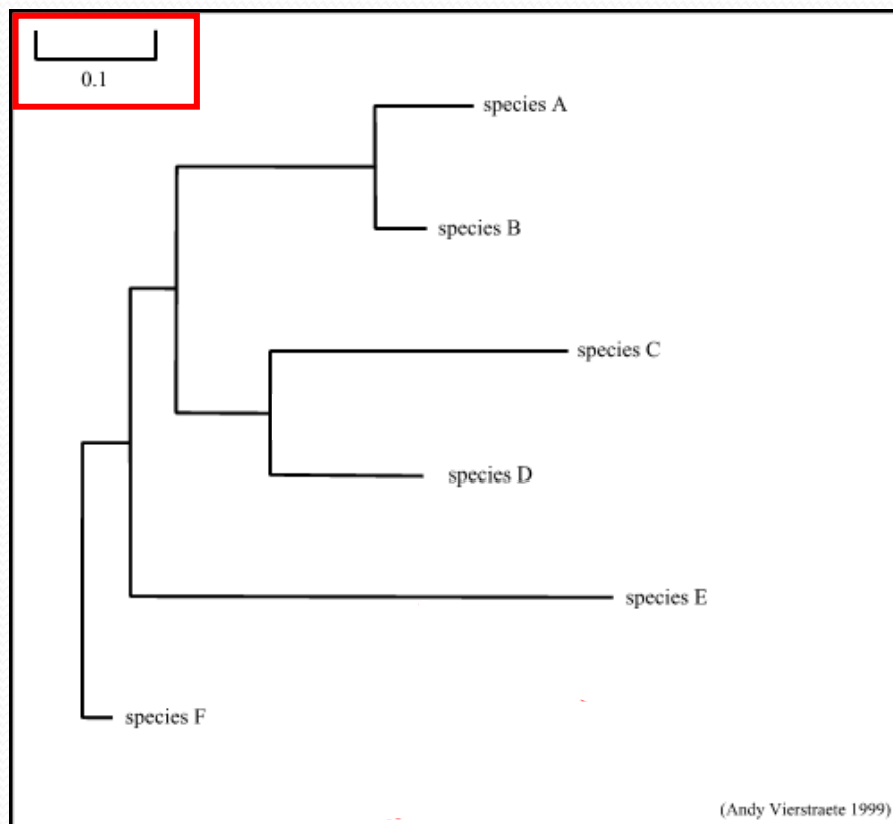
*Classical phylogenetic analysis
by Morphology*



*Molecular phylogenetic analysis
By Bio-Molecules*

Phylogenetic Tree

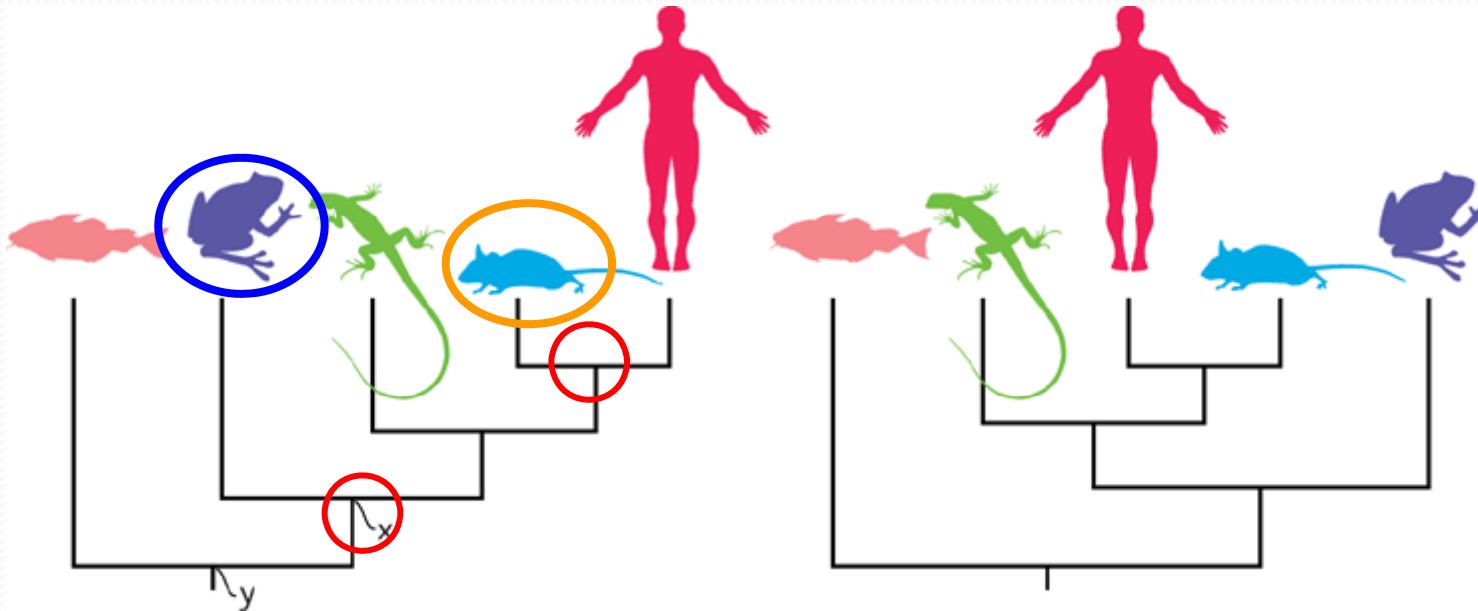
- The tree is composed of nodes connected by branches.



- **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)

Trees Only Represent The Order Of Branching

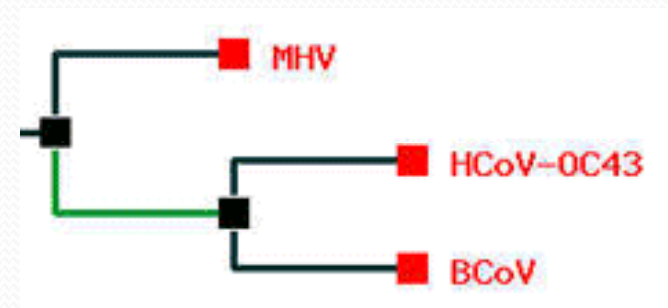
- Same topology in a different style
 - Both trees have identical topologies, with some of the internal nodes rotated.



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

The Ways to Construct the tree

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



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, consensus trees, tree distances](#)
- [Converting data, making distances or bootstrap replicates](#)

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 do 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 do 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 site 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
O          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
```

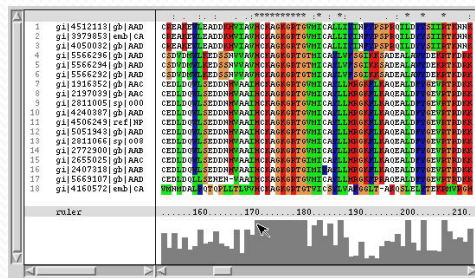
At this stage they do not have a mouse-windows interface for PHYLIP



Phylogenetic Analysis

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

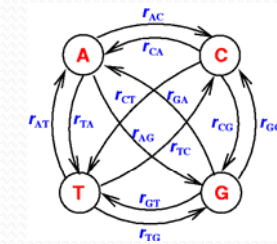
General Pipeline for Phylogenetic Analysis



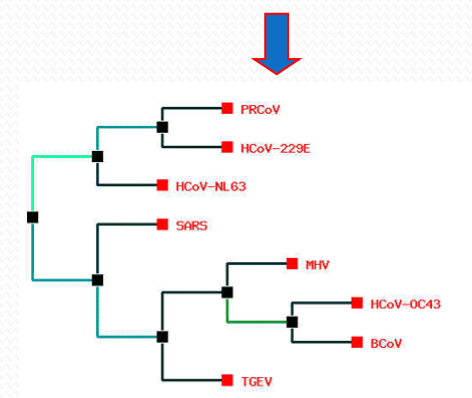
Multiple Sequence Alignment

Methods	Nucleic acid	Protein
Character state methods	<ul style="list-style-type: none"> Maximum parsimony (heuristic search) method Maximum parsimony (branch and bound search) method Compatibility method 	<ul style="list-style-type: none"> Maximum parsimony (heuristic search) method
Distance Methods	<ul style="list-style-type: none"> Distance matrix computation Neighbor-joining and UPGMA method Fitch-Margoliash and least squares method Fitch-Margoliash and least squares method with molecular clock 	<ul style="list-style-type: none"> 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 methods	<ul style="list-style-type: none"> Maximum likelihood method Maximum likelihood method with molecular clock 	

Selection of inference Methods

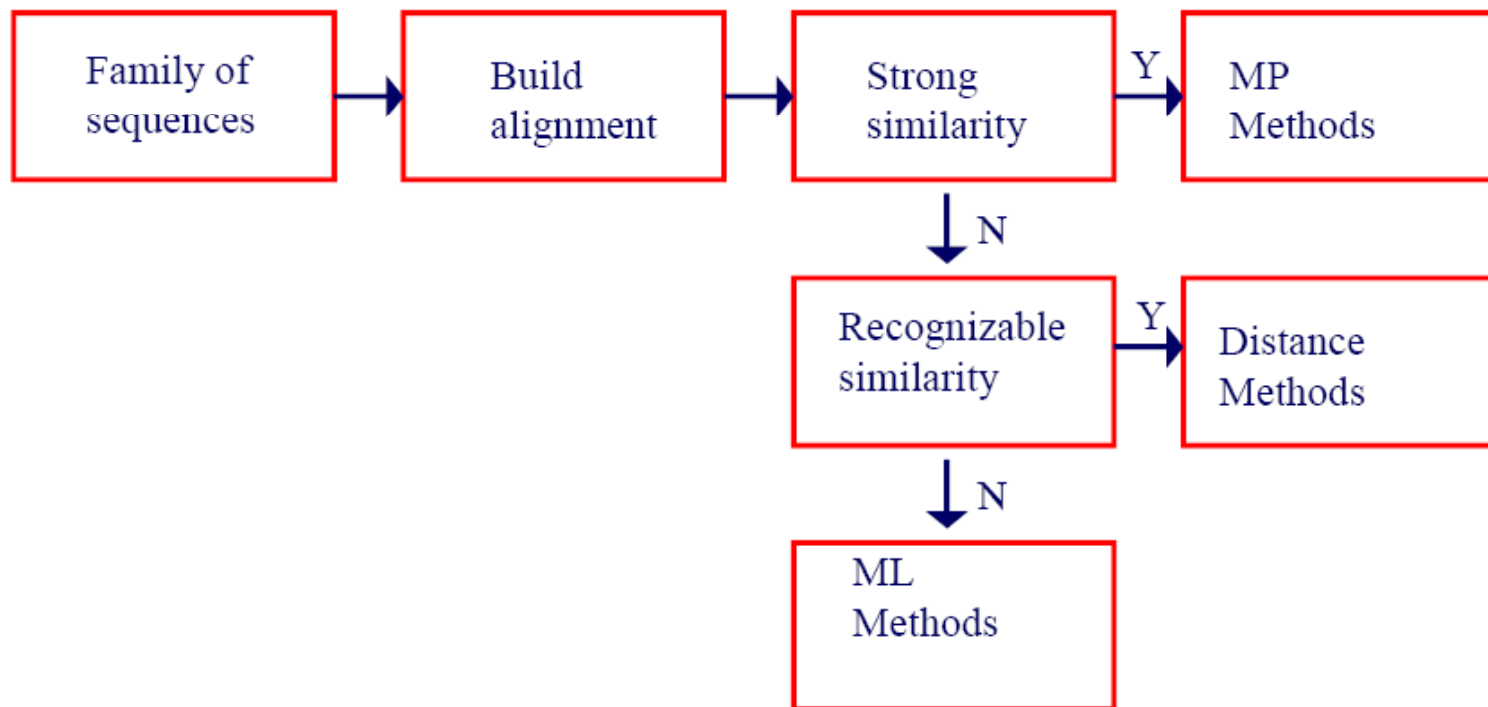


Bootstrap Substitution Model Tree Construction



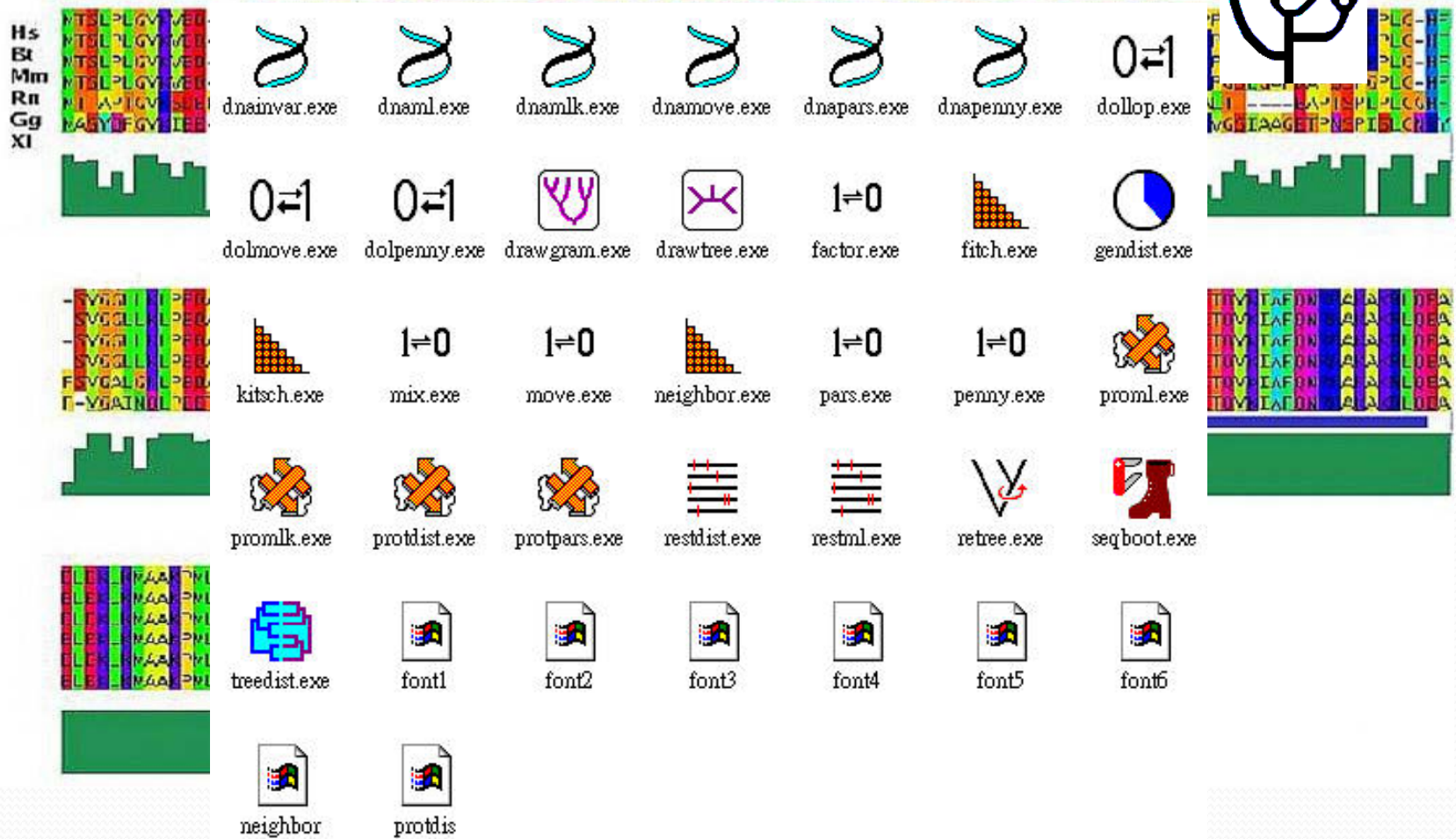
Evaluate phylogenetic tree

Flowchart of Analysis



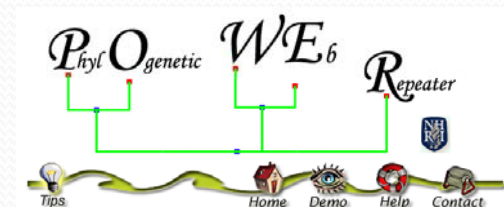
(Mount, *Bioinformatics*)

Phylogenetic Analysis Tool

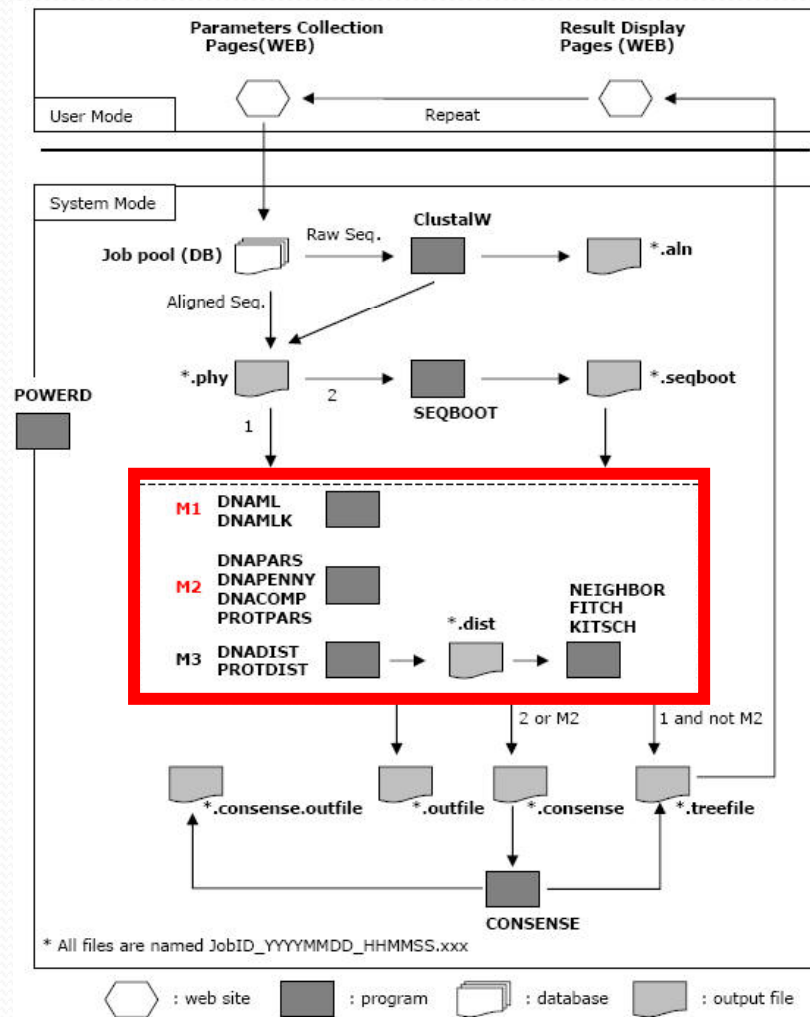


POWER: *PhylOgenetic WEb Repeater*

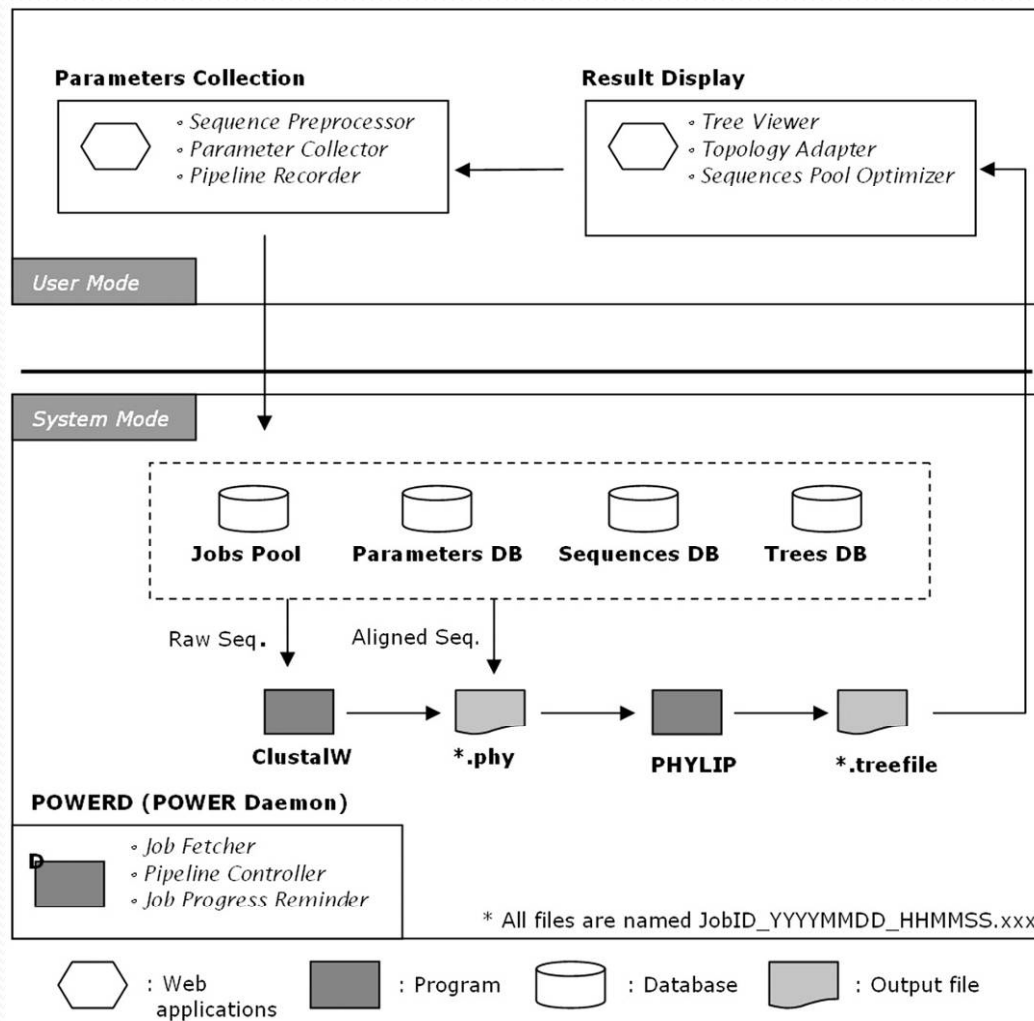
- Provide a **seamless way** to conduct the **complex phylogenetic analysis** for Biologists
- An integrated and user-optimized framework for biomolecular phylogenetic analysis
- POWER uses an open-source LAMP (Linux, Apache, MySQL, PHP) structure and infers genetic distances and phylogenetic relationships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



Integration of Phylip Packages into Automatic Flow

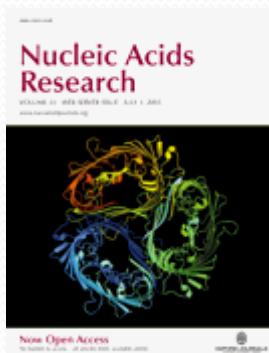


Inside of POWER



POWER: PhylOgenetic WEb Repeater

<http://power.nhri.org.tw>

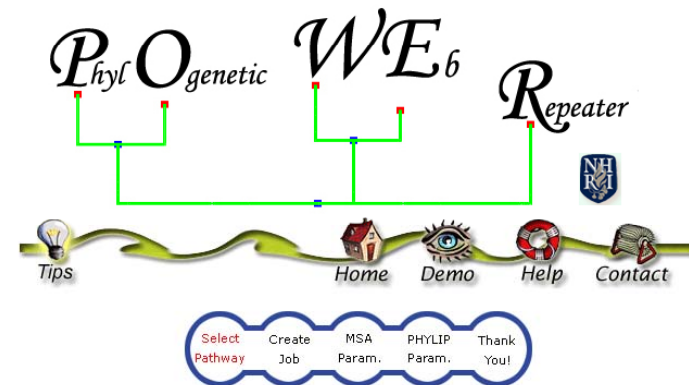


POWER: PhylOgenetic Web Repeater—an integrated and user-optimized framework for biomolecular phylogenetic analysis

Chung-Yen Lin¹, Fan-Kai Lin, Chieh-Hua Lin, Li-Wei Lai, Hsin-Jun Hsu, Shu-Hwa Chen¹ and Chao A. Hsiung

Division of Biostatistics and Bioinformatics, National Health Research Institutes 35 Keuan Road, Zhunan Town, Miaoli County 350, Taiwan ¹Institute 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 beginning 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	<input checked="" type="radio"/> MSA + Phylogenetic Analysis(Input the FASTA format) <input type="radio"/> Phylogenetic Analysis Only(Input the PHYLIP format)
Sequence Type	<input checked="" type="radio"/> DNA <input type="radio"/> Protein

PhylOgenetic Web Repeater (POWER)

Data Input

Your data type is **DNA Sequences**.

Please input your data and other related information.

Job ID* (string with character 0-9
a-z A-Z _ -)

Input sequences in FASTA format*

Example

```
seq01 2236
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
CTTGTGTTTGTGCTGATGATGATGATGATGATGATGATGATGAT
GACTGATGATGATGATGATGATGATGATGATGATGATGATGATG
TATGATGATGATGATGATGATGATGATGATGATGATGATGATG
TATGATGATGATGATGATGATGATGATGATGATGATGATGATG
seq02 1845
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
seq03 1845
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
ATGTTTAAATCTGATGAGAGATGATGATTTTGTGATTAATGACTG
```

Note:

- Length of sequence ID should be less than 10!
- Only 'A-Z', 'a-z', '0-9', '.', '-' are valid for Sequence ID!

Or load it from disk

Job Note

E-Mail (please use ',' to separate multiple email addresses)

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

NEXT

MSA parameter selection

Your data type is **DNA Sequences**.

Please select the parameters for multiple sequence alignment.

Use which algorithm for pairwise alignment

Fast-Approximate # Slow-Accurate

Pairwise alignment

Fast-Approximate algorithm

Multiple bond/size (should be integer between 1 and 4)

Topologys (should be integer between 1 and 50)

Window size (should be integer between 1 and 50)

Gap penalty (should be integer between 1 and 100)

Score type

Multiple alignment

Gap opening penalty (should be real between 0.0 and 100.0)

Gap extension penalty (should be real between 0.0 and 10.0)

ChkA weight matrix (1-2)

Transition weighting (should be real between 0.0 and 1.0)

%Identical for delay (should be integer between 0 and 100)

NEXT

Phylogeny inference

Your data type is **DNA**.

Please select the method for phylogenetic analysis.

Character state methods

- [Maximum parsimony\(heuristic search\) method](#)
- [Maximum parsimony\(branch and bound search\) method](#)
- [Compatibility method](#)

Distance Methods

- [Neighbor-joining and UPGMA method](#)
- [Fitch-Margolash and least squares method](#)
- [Fitch-Margolash and least squares method with molecular clock](#)

Maximum likelihood methods

- [Maximum likelihood method](#)
- [Maximum likelihood method with molecular clock](#)

BACK

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	<input type="text" value="777"/> (must be odd)
Number of replicates	<input type="text" value="100"/>
Resampling methods	<input type="text" value="Bootstrap"/>

[BACK](#) [NEXT](#)

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	<input type="text" value="Kimura 2 parameters"/>
Transition/transversion ratio	<input type="text" value="2"/> (must be a positive real number)

[BACK](#) [NEXT](#)

Selected method for phylogeny inference

Your data type is DNA Sequences.

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

Tree constructing method	<input type="text" value="Neighbor-joining"/>
Outgroup root	<input type="text" value="0"/> (the species being taken in the numerical order that they occur in the input file)
Randomize input order of species	<input type="radio"/> Yes <input checked="" type="radio"/> No

[BACK](#) [NEXT](#)

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 16:06:26

Dear Sir or Madam:

The job 'comonavirus0720' you sent at '2004-07-20 12:11:37' has finished!
The whole process that started at 2004-07-20 18:06:13 and finished at 2004-07-20 18:03:35 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:
Deposition:
http://211.76.166.77/power/result_page.php?job_no=2041&job_name=comonavirus0720_0720_170017

POWER version 1.0
PhyLP package version 3.5
ClustalW version 1.02

May the POWER with you!

WARNING
System will **CLEAN** job data regularly!

We recommend you to **SAVE** the phylogenetic tree image yourself.
(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 for creating new job. Click and get it back.

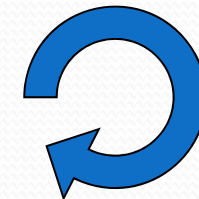
[TREE PARAMETER]
 X factor: [1] Y factor: [1]

>> JOB INFORMATION

[Job Parameters]			
Job ID	comonavirus0720	Sequence Type	DNA
Job Note	illustration		
[ClustalW Parameters]			
kupile	2	topdiag	4
window_size	4	-cigap	5
scorey_w	FRZE V	o_gapopen	15
pwgapxt	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
[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	comonavir_0720_0720_170017
TREE IMAGE	comonavir_0720_0720_170017_3306.png
CLUSTALW	comonavirus0720_0720_170017aln
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* (string with character 0-9
a-z A-Z _ -)

Input sequences in FASTA format*

Note:
- Length of sequence ID should be less than 10!
- Only 'A-Z', 'a-z', '0-9', '-', '_' are valid for Sequence ID!

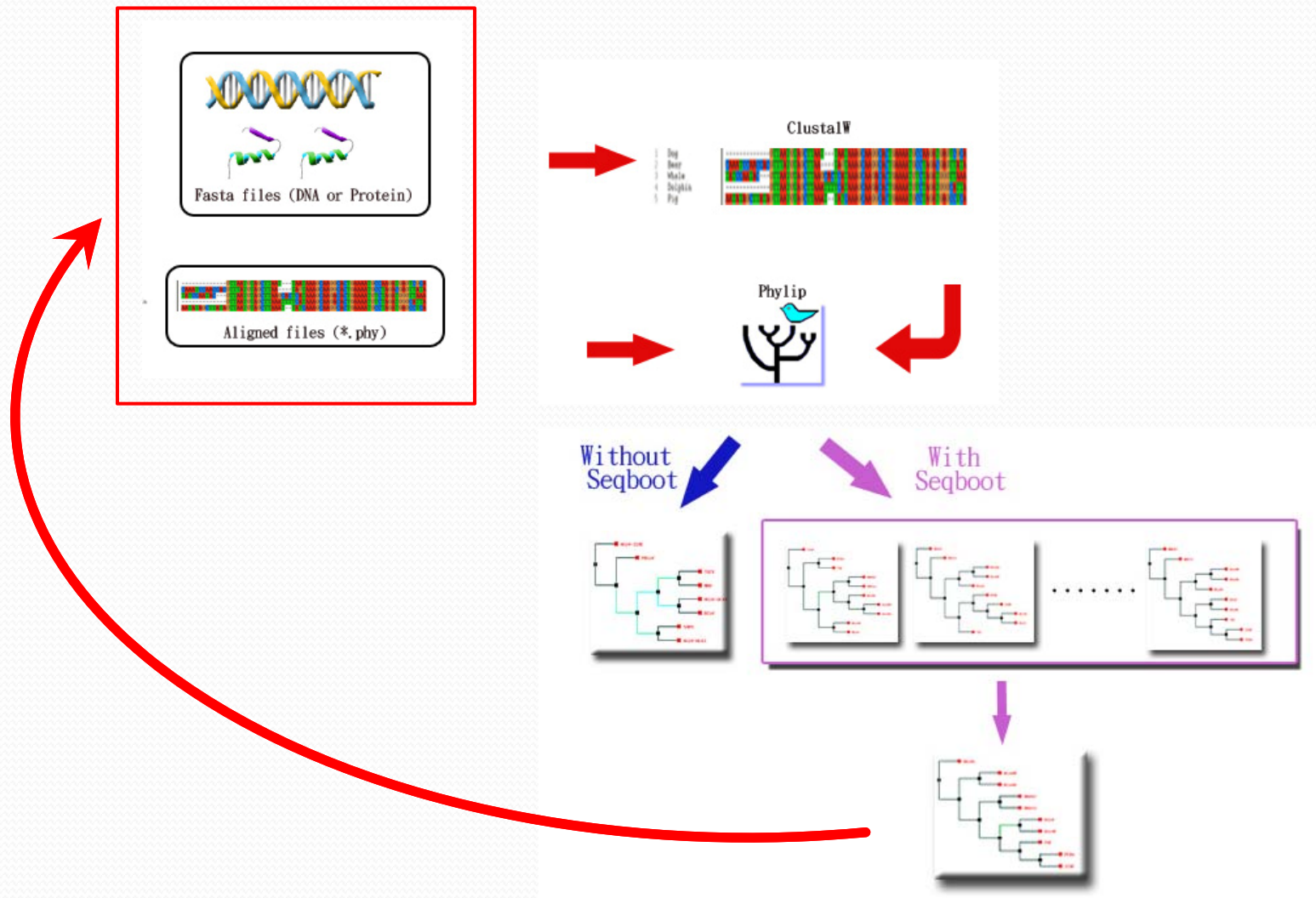
Or load it from disk

Job Note

E-Mail
(please use ',' to separate multiple email addresses)

Re-perform the process by items added or deleted

PhylOgenetic Web Repeater (POWER)



Add/ delete sequences to invoke new job

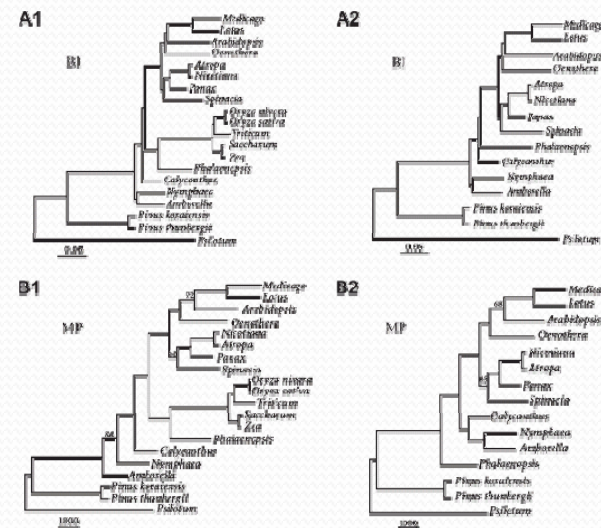
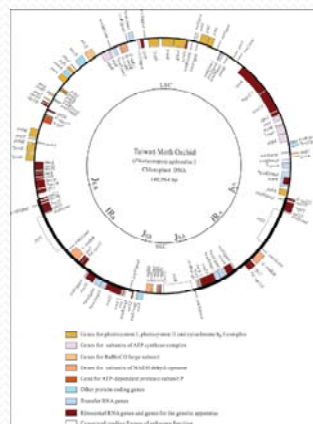
Publication in *POWER*

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

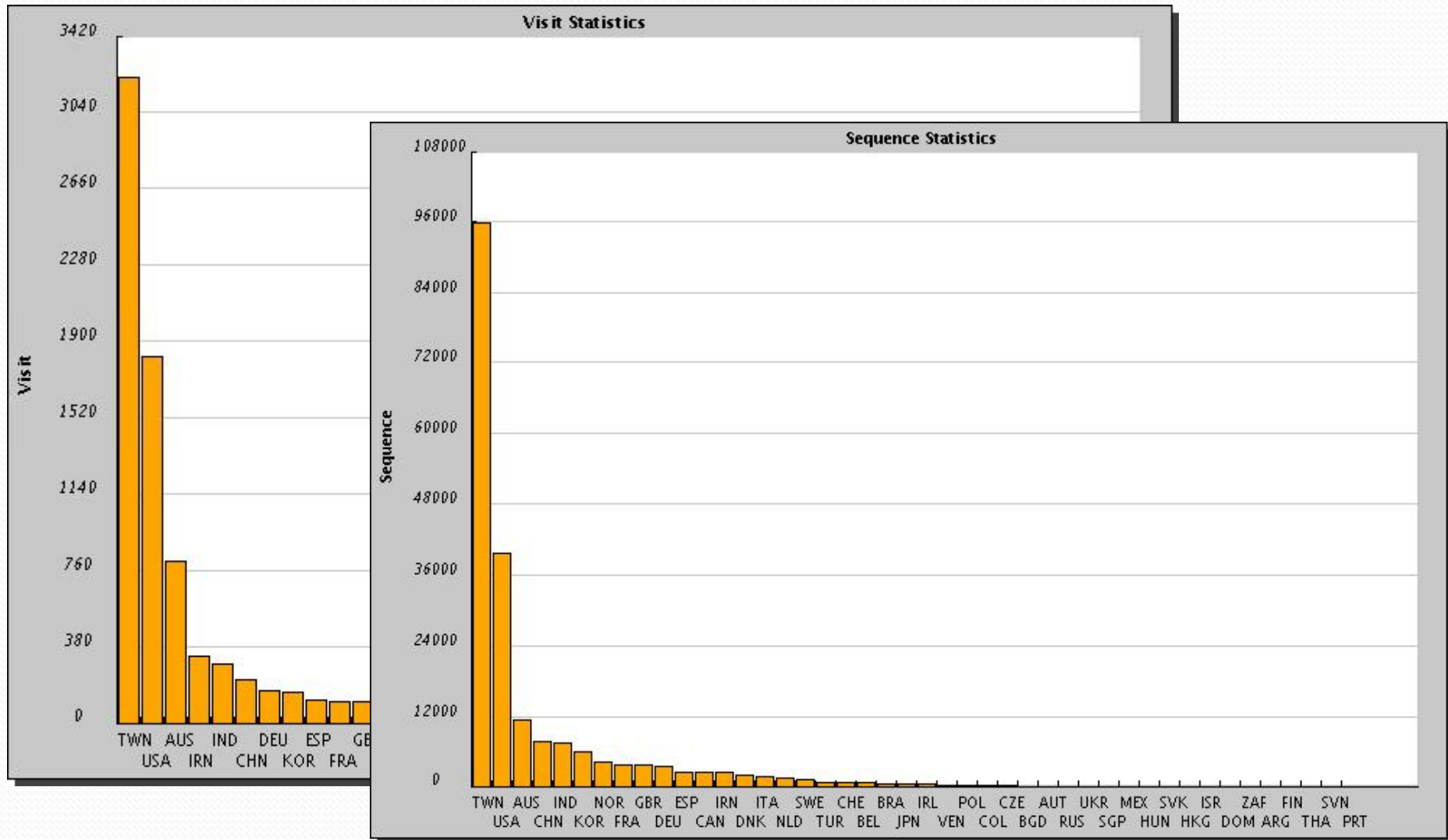
Mol. Biol. Evol. 23(2):279–291. 2006

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

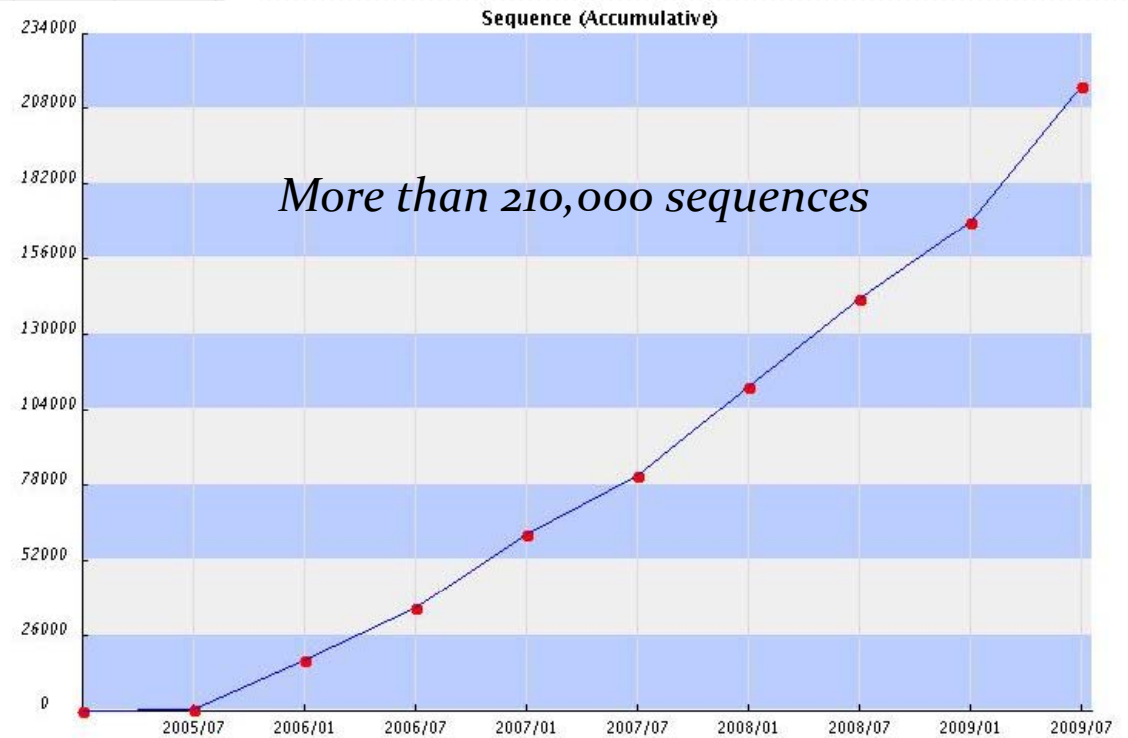
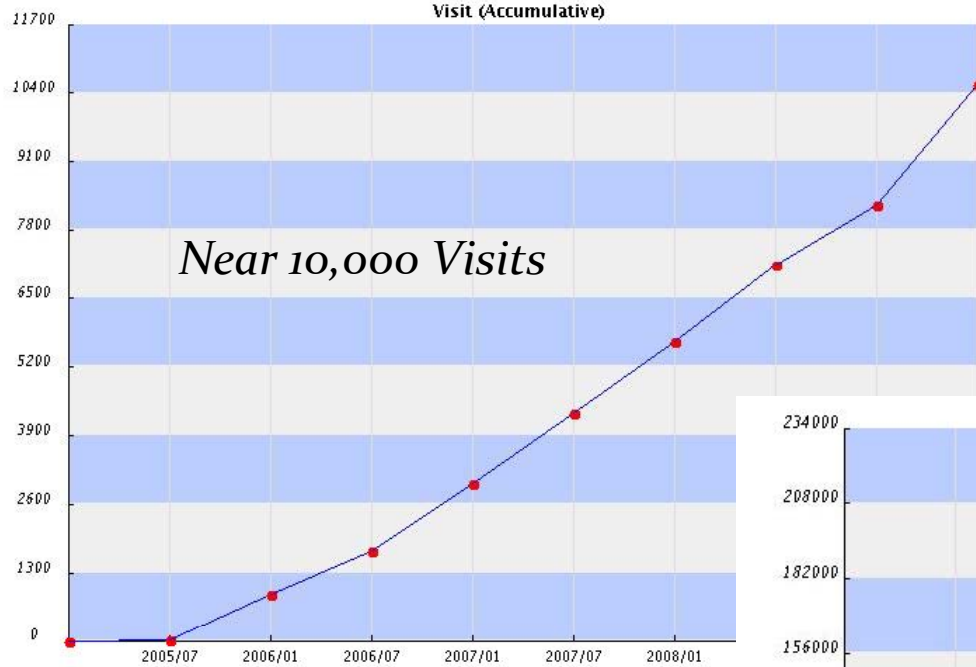
*Institute of Biotechnology, National Cheng Kung University, Tainan, Taiwan; †Department of Superintendent, Tainan Municipal Hospital, Tainan, Taiwan; ‡Institute of Plant and Microbial Biology, Academia Sinica, Taipei, Taiwan; §Department of Life Sciences, National University of Kaohsiung, Kaohsiung,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



Service Usage of POWER from 2005 July.



Service Usage of POWER from 2005 July.



Automatic On-Line Demonstration

The PhylOgenetic Web Repeater (POWER) allows users performing phylogenetic analysis with molecular data by most programs of PHYLIP package repeatedly. POWER provide two pipelines to process the analysis. One of them includes multiple sequence alignment (MSA) at the beginning 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	<input type="radio"/> MSA + Phylogenetic Analysis(Input the FASTA format) <input type="radio"/> Phylogenetic Analysis Only(Input the PHYLIP format)
Sequence Type	<input type="radio"/> DNA <input type="radio"/> Protein

NEXT

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

Research article

Open Access

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

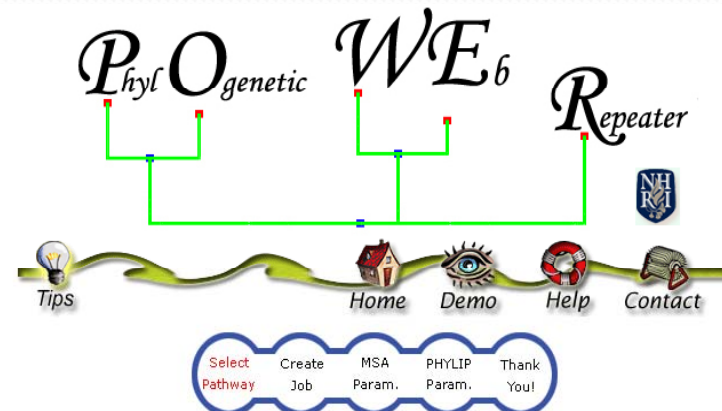
César L Avila¹, Viviana A Rapisarda¹, Ricardo N Farías¹, Javier De Las Rivas² and Rosana Chehín*¹

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

Correspondence: César L Avila - avila@iqb.unqf.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>]

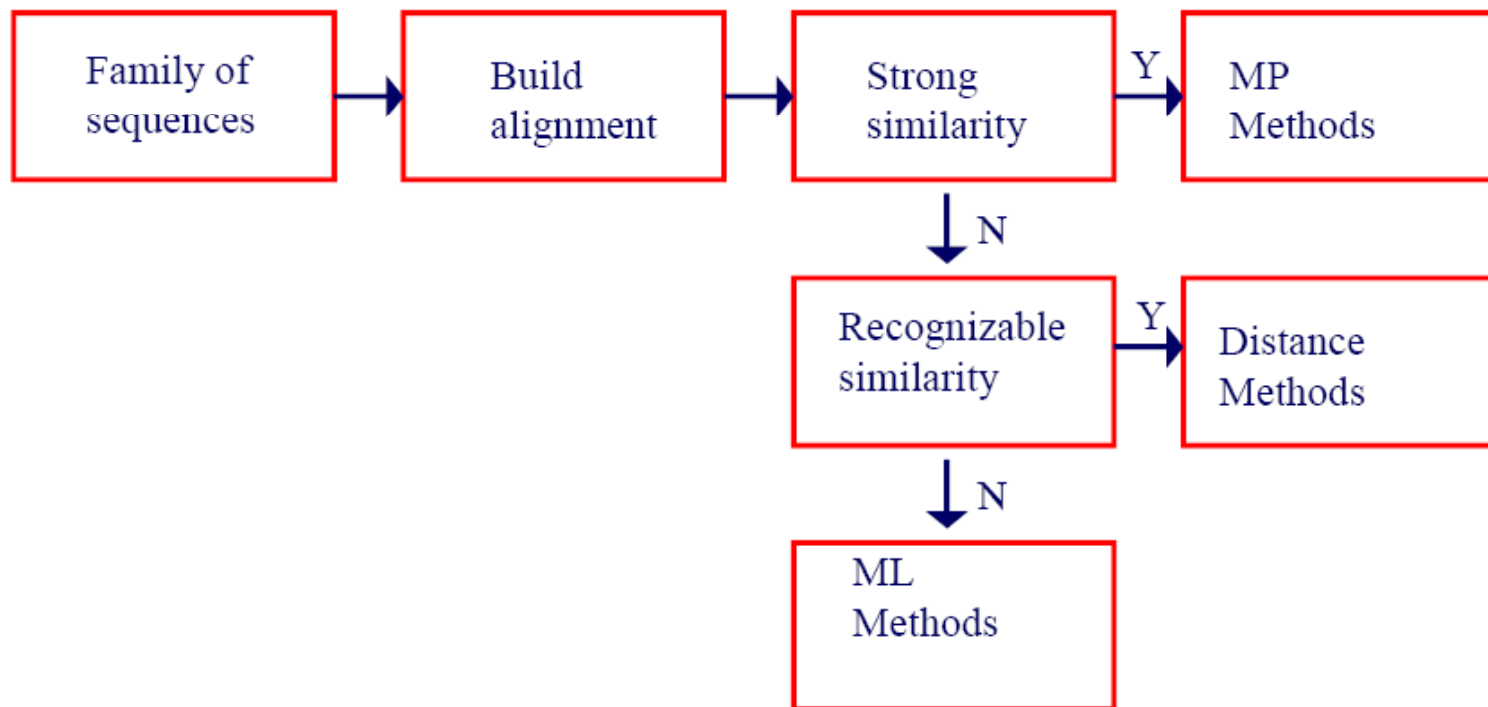
POWER Listed in

- 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



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

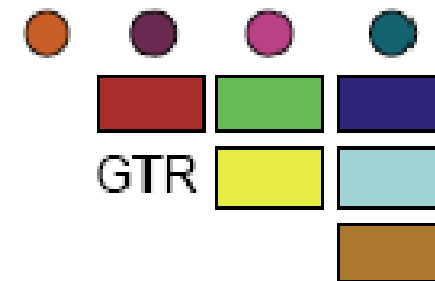
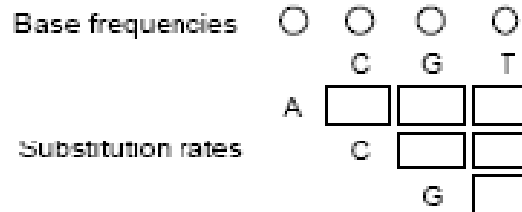
$$L(\tau, \theta) = \text{Prob}(\text{Data} | \tau, \theta)$$

Or

$\text{Prob}(\text{Aligned Sequences} | \text{tree, model of evolution})$

Illustration of DNA substitution Model

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



GTR (for four characters, as is often the case in phylogenetics) requires 6 substitution rate parameters ($x_1 \sim x_6$), as well as 4 equilibrium base frequency parameters.

Relationships Among Some Standard Models Of Nucleotide Evolution

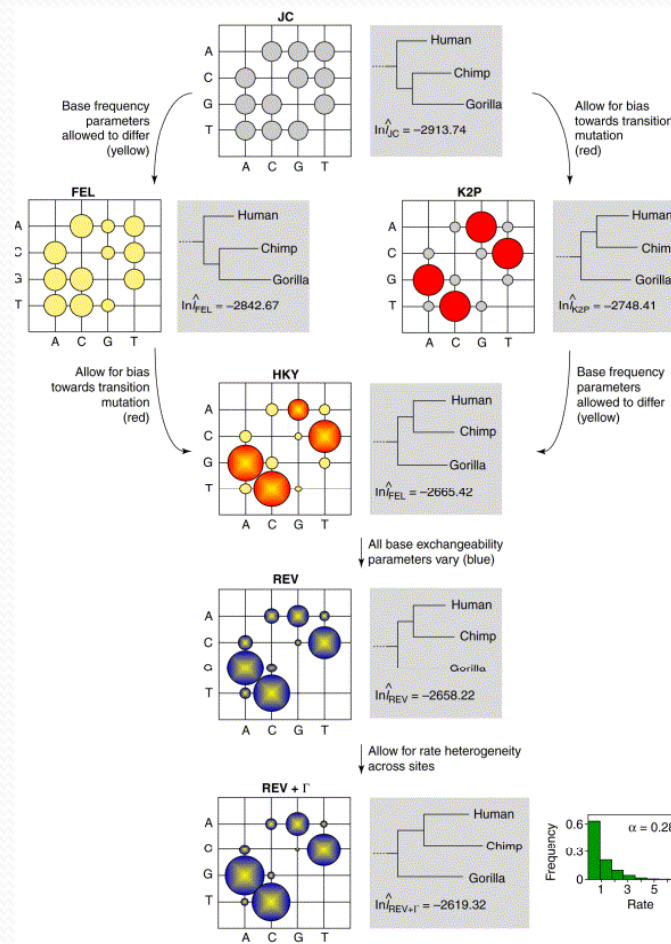
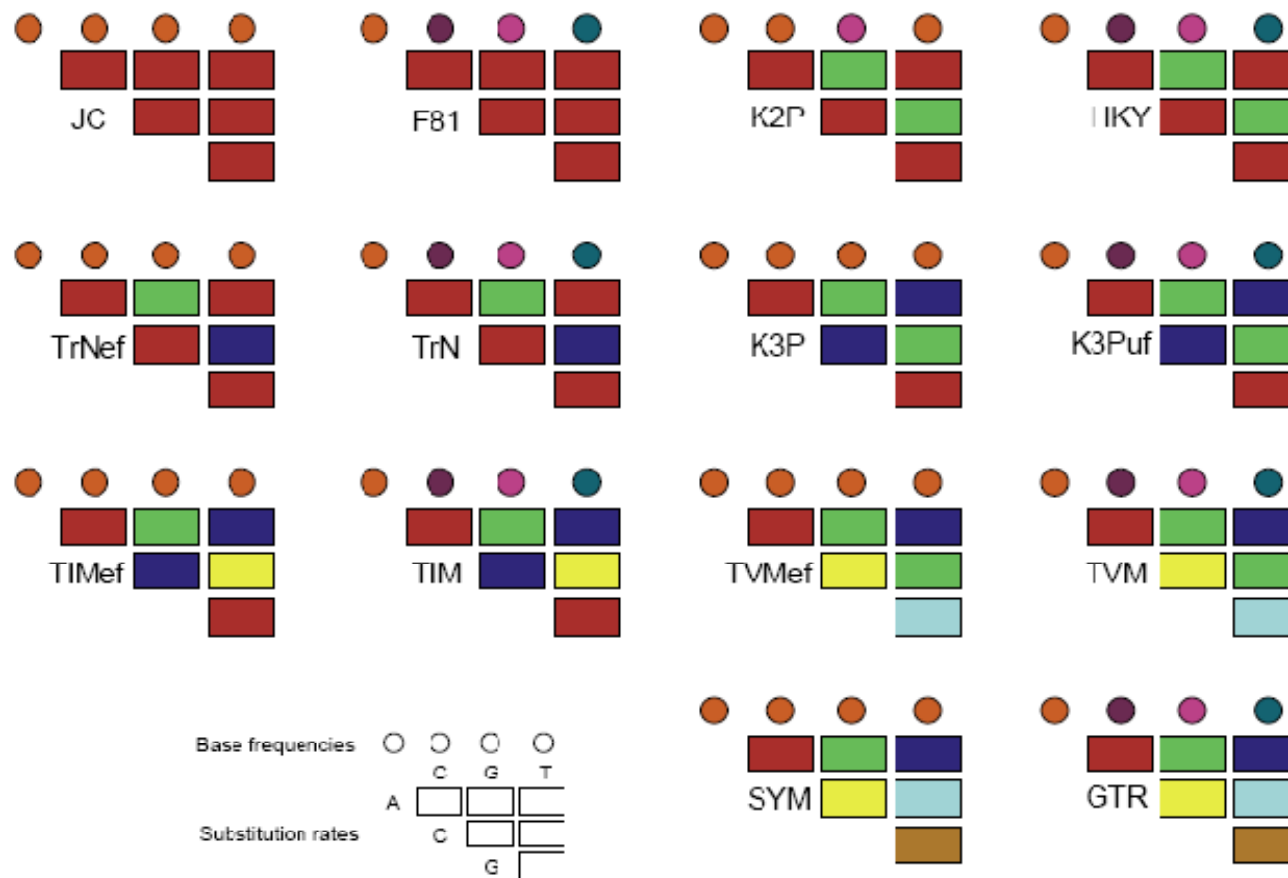
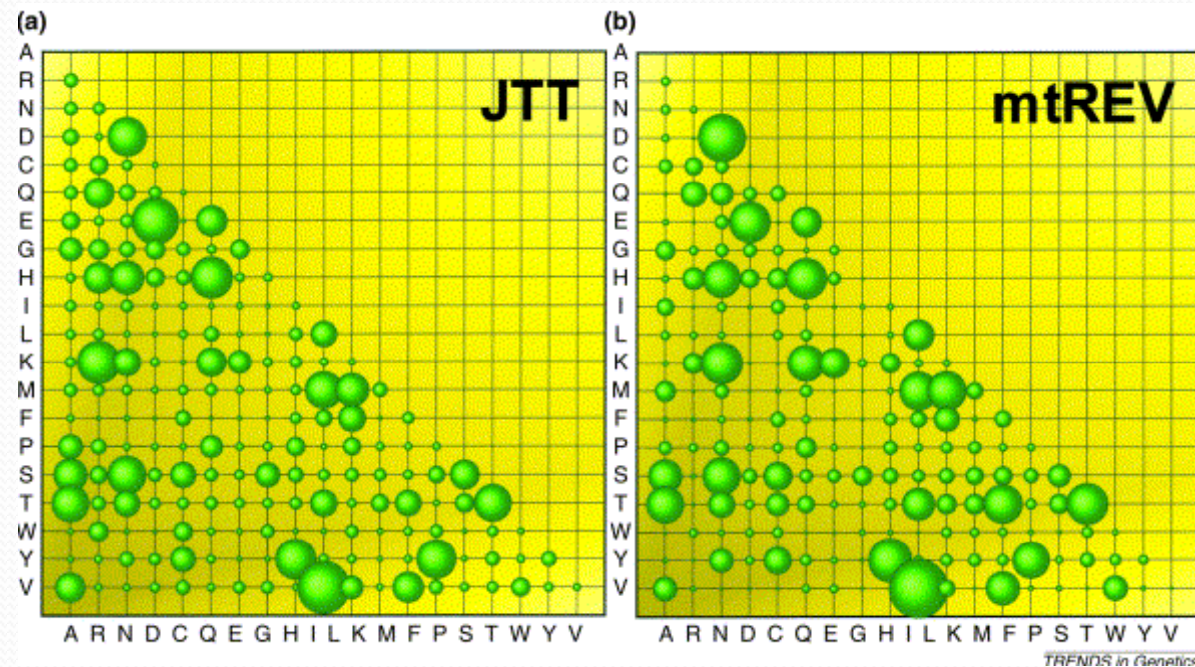


Illustration of Models for DNA



Models of Amino Acid Replacement



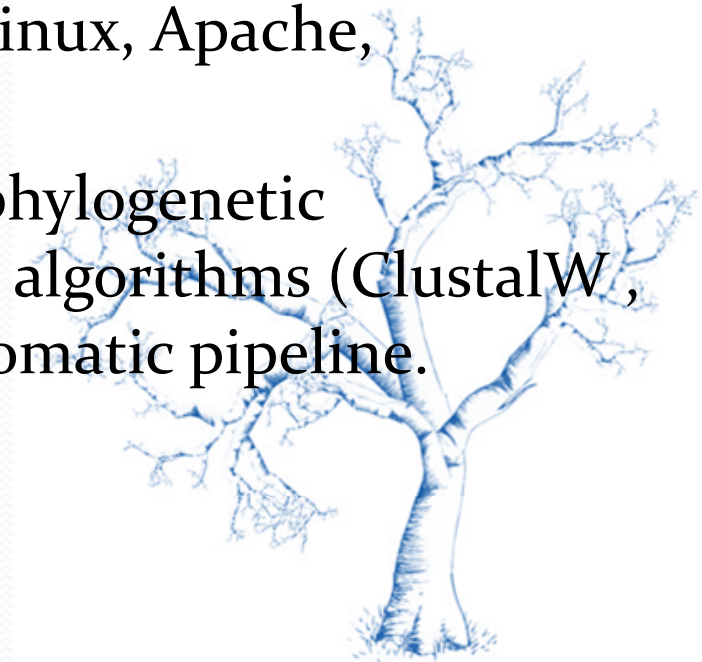


Background for PALM

- 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.

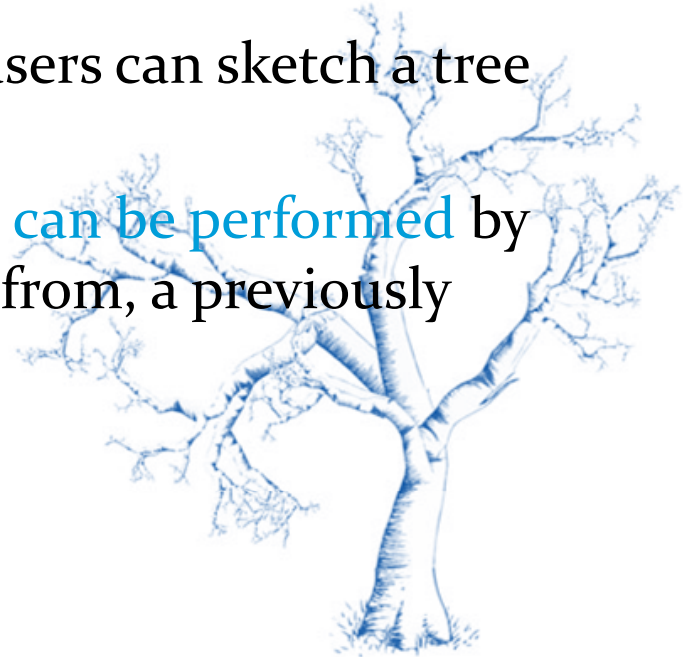
Motivation I

- 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 2.0
- ClustalW 2.0.8
- Seqret (EMBOSS)



Phylogenetic reconstruction by Automatic Likelihood Model selector

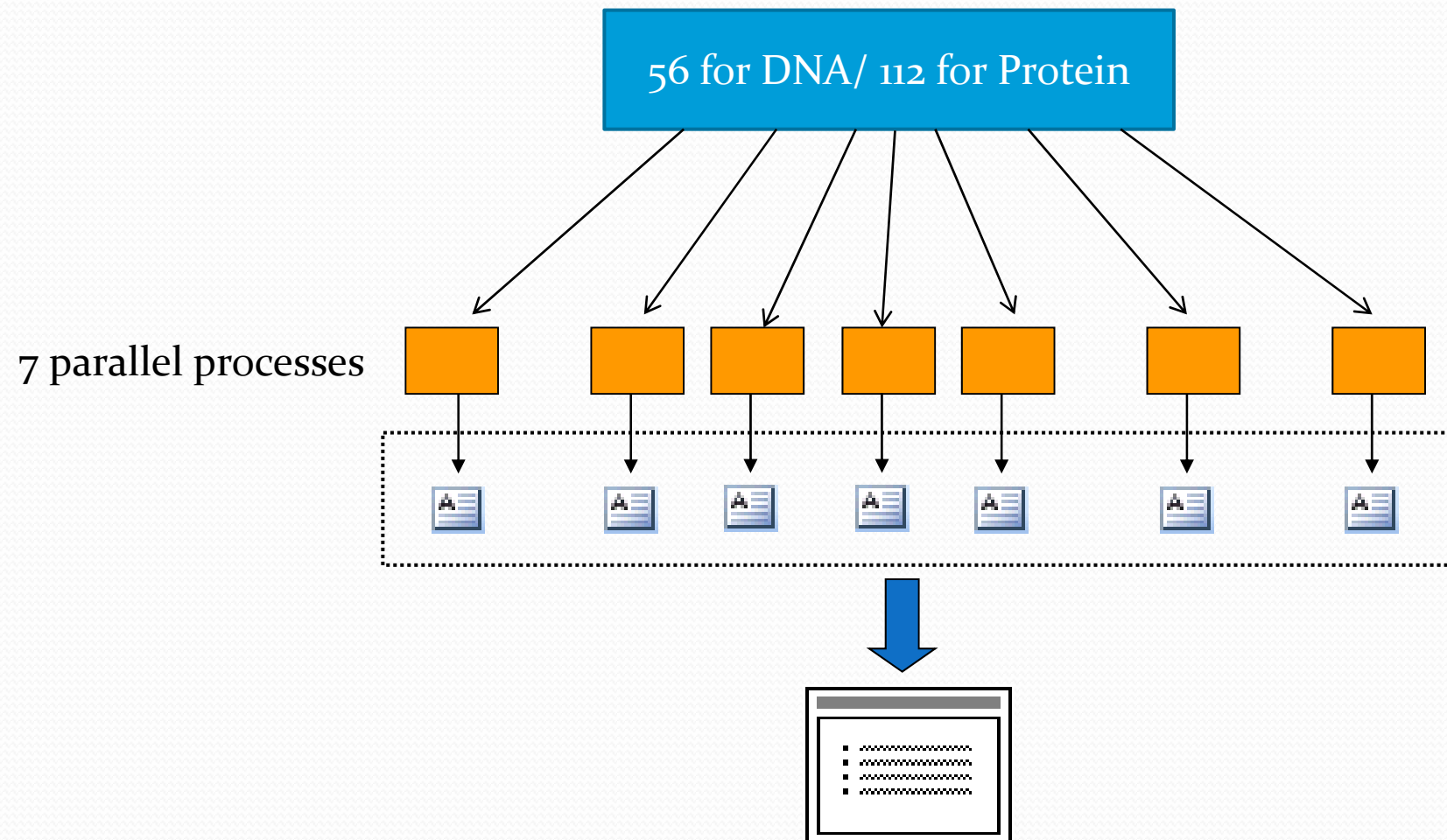
Home Demo Help Contact

Input Sequences	
Input type	<input type="radio"/> Sequence in FASTA format <input type="radio"/> Aligned sequence in PHYLIP format
Sequence type	<input type="radio"/> DNA <input type="radio"/> Protein
Sequences*	<div style="border: 1px solid #ccc; height: 60px; width: 100%;"></div> <p>Clear Input</p> <input type="text"/> <input type="button" value="浏览..."/> <input type="checkbox"/> example file
Number of bootstrap data sets	100 <input type="button" value="v"/> <input type="checkbox"/> Print bootstrap information
Job Note	<input type="text"/>
Enter your email*	<input type="text"/>
▼ Advanced Option	
Number of substitution rate categories	4 <input type="button" value="v"/>

Models Used in PALM

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

Distribution Computing by PalmMonitor for Each Substitution Model



Decreasing Time by PALMmonitor

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

- JTT, MtREV 1h:22:21
- MtMam, MtArt 1h:51:25
- Dayhoff ,WAG 1h:33:51
- RtREV, CpREV 1h:33:50
- Blosum62, VT 1h:15:58
- HIVb, HIVw 1h:35:02
- LG, DC, Mut 1h:38:36



Over 10 Hours

- All Models by PALMmonitor

1h:11:11

Source: 99 sequences with 247 residues for each

Parallel Computing on Bootstrapping

DNA : [DNA_Big_24.phy](#) (24 sequences, average 5000 bps substitution model: HKY85- Default)

Bootstrap	100	1000
Runtime	11h:47m:5s	~120 h
Runtime (5 cores/ 8 cores)	2h:53m:2s	17h:52m:15s



1/6 ↓

Protein : [Pseq](#) (20 sequences, average 820 a.a., substitution model: LG - Default)

Bootstrap	100	1000
Runtime	17h:31m:33s	~175 h
Runtime (5 cores/ 8 cores)	5h:15m:19s	36h:55m:10s

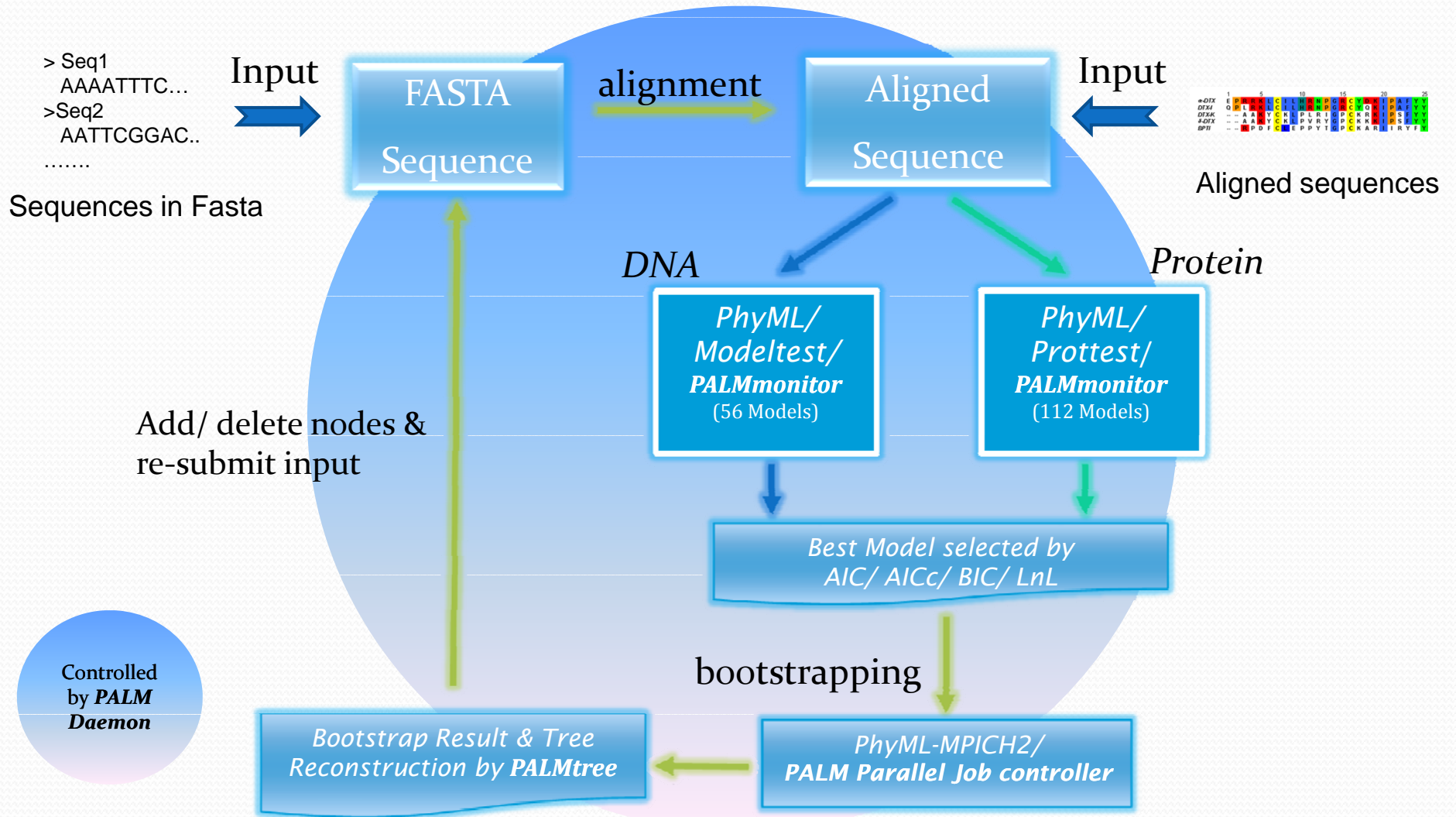
1/6 ↓



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 selected by PALM

Flowchart of PALM



Result of PALM

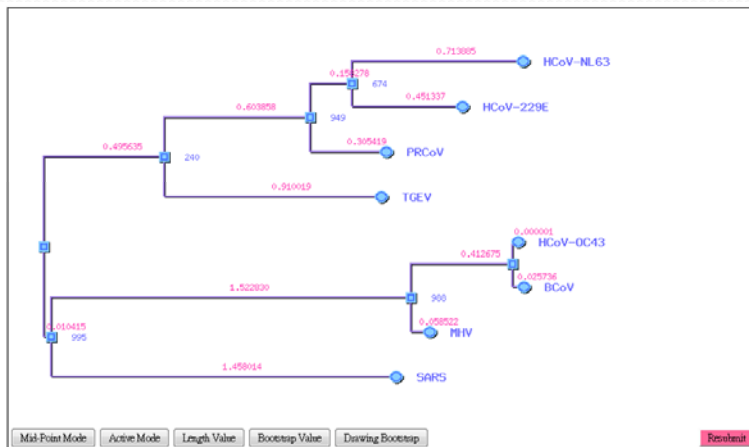
A



PALM Result

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

B



C

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

D

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

E

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

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

Demonstration of PALM



Input Sequence

Input Type Sequence in FASTA format
 Aligned sequence in PHYLIP format

Sequence Type DNA Protein

Example File

Sequences

Clear Input

Number of Bootstrap Data Sets: 100

Job Note

Enter Your Email*

*: email is optional. Users also can receive the notification immediately via mail when jobs are done.

Advanced Option

Number of Substitution Rate Categories: 4

Starting Tree (Newick Format) Build BioNJ tree User tree

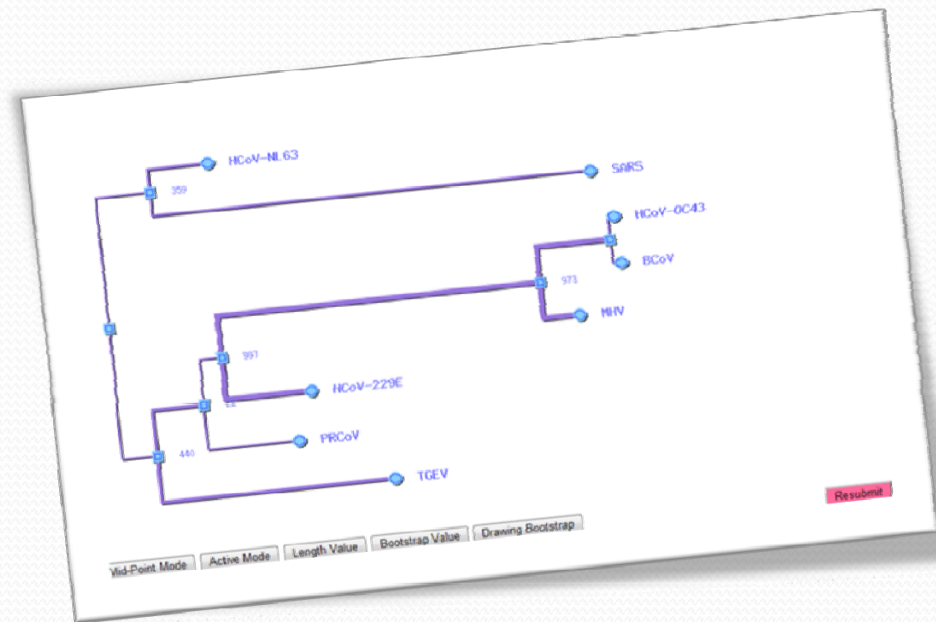
Model Selection Criterion: AIC

Optimize Tree Topology and Branch Lengths Yes No

Submit Reset

Current Status in Queue: 1 job(s) remaining in the queue.

© 2008 Systems Biology & Network Biology Lab,



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

Demo Flash of PALM



Demo (Please click the following vedio clip)

English 中文

1. Create a Job



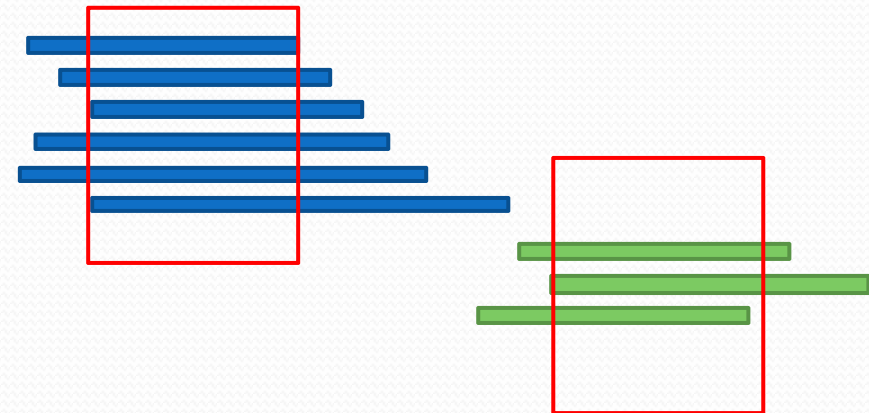
Users can paste their specific and interesting sequences in the below area, or select the Example File with related options.

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

Some Suggestions

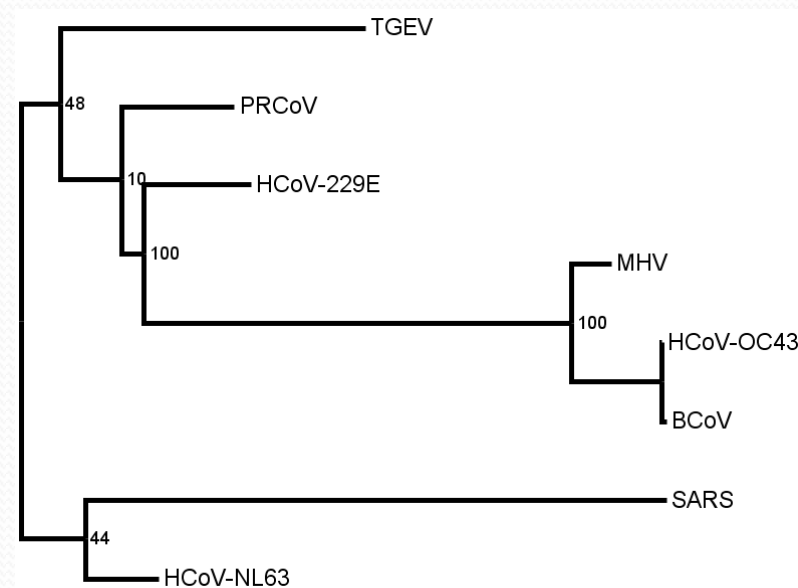
- Please be careful to choose the sequences
- Only well aligned sequences lead to meaningful phylogenetic result.
- RNA editing may introduce bias during analysis. Avoid those regions that may have such conditions.

```
Q5E940_BOVIN -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_BUMBA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_MOUSE -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_RAT -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_CECIL -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_RABST -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
Q72UCU_BURSE -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_ICTPO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_DROME -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_DICDI -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
Q5ALPQ_DICDI -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PILAT -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_SULAC -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_DULTO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_DULDO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_AEHPK -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PTRAN -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METAC -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METHA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_ACTFO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METKA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METTE -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METTL -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METVA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_METJA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PTRAD -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PTRHO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PTRFU -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PTRKO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_HALMA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_HALVO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_HALSA -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_THREAC -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_THREVD -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
RLAA_PICTO -----MREEDRTWRSWVLEKLELDGQCFIVGADVSRGQIISLQKAVVLMGKSTMRKAIKRLKRW---FALE 74
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90
```



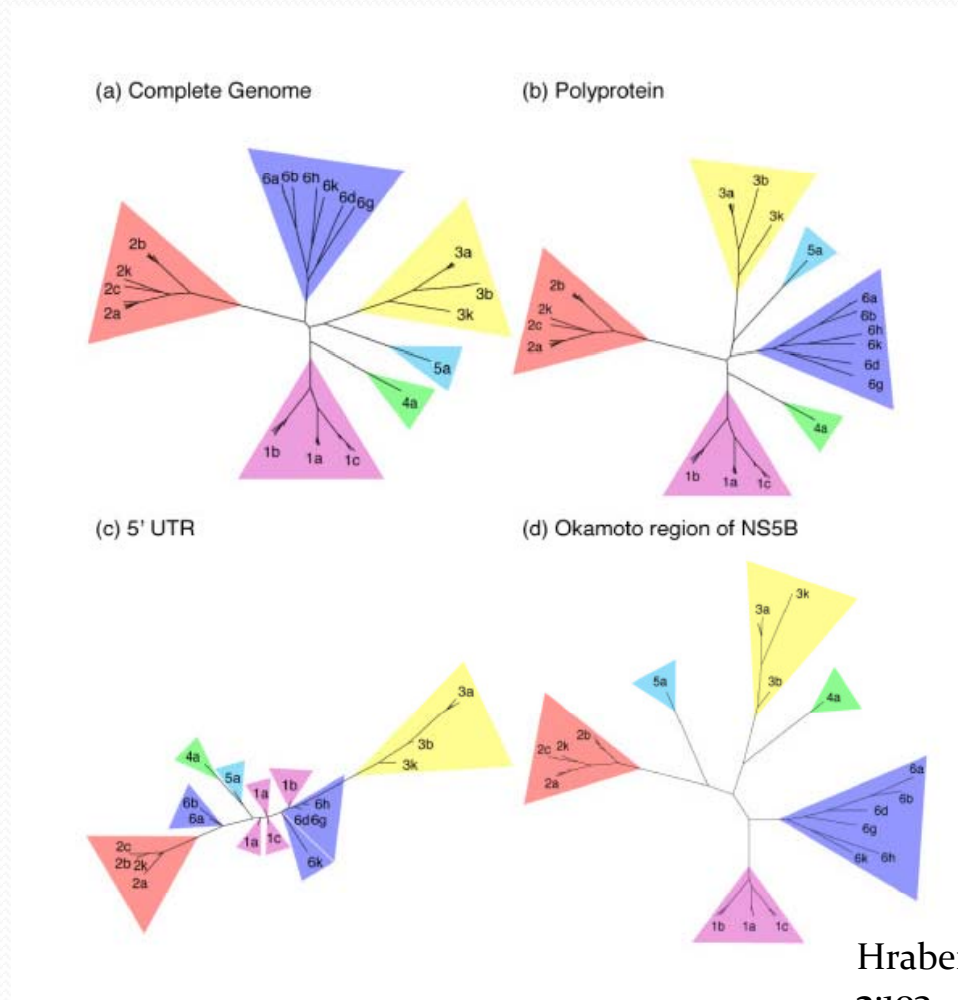
Bootstrap (BS) Analysis

- 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 cannot be corrected by the re-sampling process.



Input Sequences Make the Tree Different

HIV



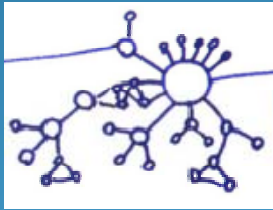
Hraber *et al.* *Virology Journal* 2006
3:103 doi:10.1186/1743-422X-3-103

Future Plans for PALM

- Integrate more substitution models into PALM
- Improve and optimize the performance of whole pipeline
- MrBayes will be implemented into this system for Bayesian inference.
- Parallel computing for large scale, ie. 16S RNA tree (near 2000-4000 sequences) reconstruction in metagenomics for revealing microbial community




Interesting Web Applications and On-going Projects



系統生物與網路生物學實驗室
Laboratory of Systems Biology and Network Biology


MyBLAST (Customized BLAST Framework)

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



Email

PWD

 [Has not registered?](#)
[Login](#) [Forget password](#)

- [DB Management](#)
- [Upload DB](#)
- [Run BLAST](#)
- [View Results](#)
- [User Guide](#)










:: 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 ::

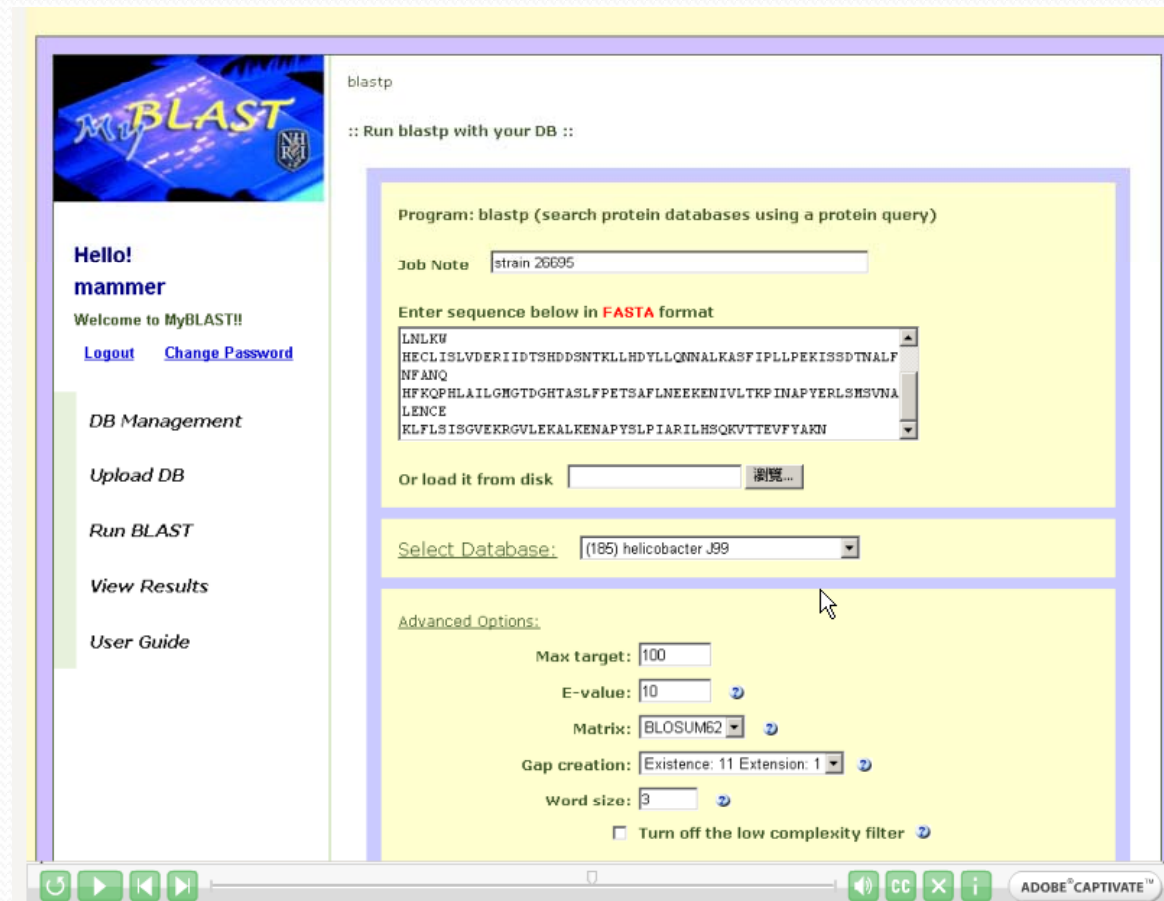
DB description: (6) HP 26695 Matche Sequences

Submit description: (17) J99 blast 26695 Top 3

Download Output Files ([text file](#)) or ([csv file](#))

Seq.	Rank	Hits	E-value	Score	Bits	Match Length	Identities
g 15611072 ref NP_222723.1 transcription antiterminatorprotein NusB [Helicobacter pylori 399]	1	g 15644635 ref NP_206803.1 transcription antitermination protein NusB [Helicobacter pylori 26695]	0.0	267.0	583	138	136/137 (99%) Positives = 136/137 (99%)
g 15611073 ref NP_222724.1 riboflavin synthase subunit beta[Helicobacter pylori 399]	1	g 15644636 ref NP_206804.1 riboflavin synthase subunit beta [Helicobacter pylori 26695]	0.0	294.0	733	156	147/155 (94%) Positives = 152/155 (98%)

Flash Demo for MyBLAST



The screenshot displays the MyBLAST web interface. On the left, a sidebar contains a logo, a greeting "Hello! mammer", a welcome message "Welcome to MyBLAST!", and links for "Logout" and "Change Password". Below these are navigation links for "DB Management", "Upload DB", "Run BLAST", "View Results", and "User Guide". The main content area is titled "blastp" and includes a note: ":: Run blastp with your DB ::". It features a "Program: blastp (search protein databases using a protein query)" section with a "Job Note" field containing "strain 26695". A text area for the sequence is pre-filled with a FASTA format protein sequence. Below this is an "Or load it from disk" option with a file selection button. The "Select Database:" dropdown is set to "(185) helicobacter J99". An "Advanced Options:" section includes fields for "Max target: 100", "E-value: 10", "Matrix: BLOSUM62", "Gap creation: Existence: 11 Extension: 1", and "Word size: 3". A checkbox for "Turn off the low complexity filter" is present and unchecked. The interface is framed by a yellow border and includes a video player control bar at the bottom with a "ADOBE CAPTIVATE" logo.

blastp

:: Run blastp with your DB ::

Program: blastp (search protein databases using a protein query)

Job Note: strain 26695

Enter sequence below in FASTA format

```
LNLKW
HECLISLVDERIIDTSHDDSNTKLLHDYLLQNNALKASFIPLLPEKISSDTNALF
NFANQ
HFRQPHLAILGHGTDGHTASLFPETS AFLNEEKENIVLTKP INAPYERLSHVSUNA
LENCE
KLFLSISGVEKRGVLEKALKENAPYSLP IARILHSQKVTTEVFYAKN
```

Or load it from disk: 浏览...

Select Database: (185) helicobacter J99

Advanced Options:

Max target: 100

E-value: 10

Matrix: BLOSUM62

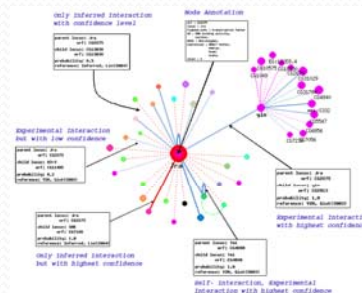
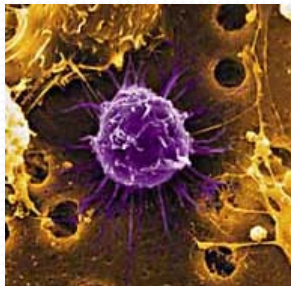
Gap creation: Existence: 11 Extension: 1

Word size: 3

Turn off the low complexity filter

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)



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

4月 2008

專案主題

- Post new topic.
- project
- 專門實驗室的合作
- ELN 電子實驗室記錄 (Electronic Laboratory Notebook) 4 1 篇文章
- MicroArray 3 1 篇文章
- Palm 1
- Molas 0
- LPS 0
- Others 2 2 篇文章

會議記錄

- Lab Meeting 4
- Group Meeting 3 3 篇文章

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

7月 2008

預覽回應

test

ELN Versions in Future



- Group version

- Linux /windows –based with fixed IP
- For laboratory/ large group
- International cooperation
- Need better hardware with few IT manpower
- Privilege control

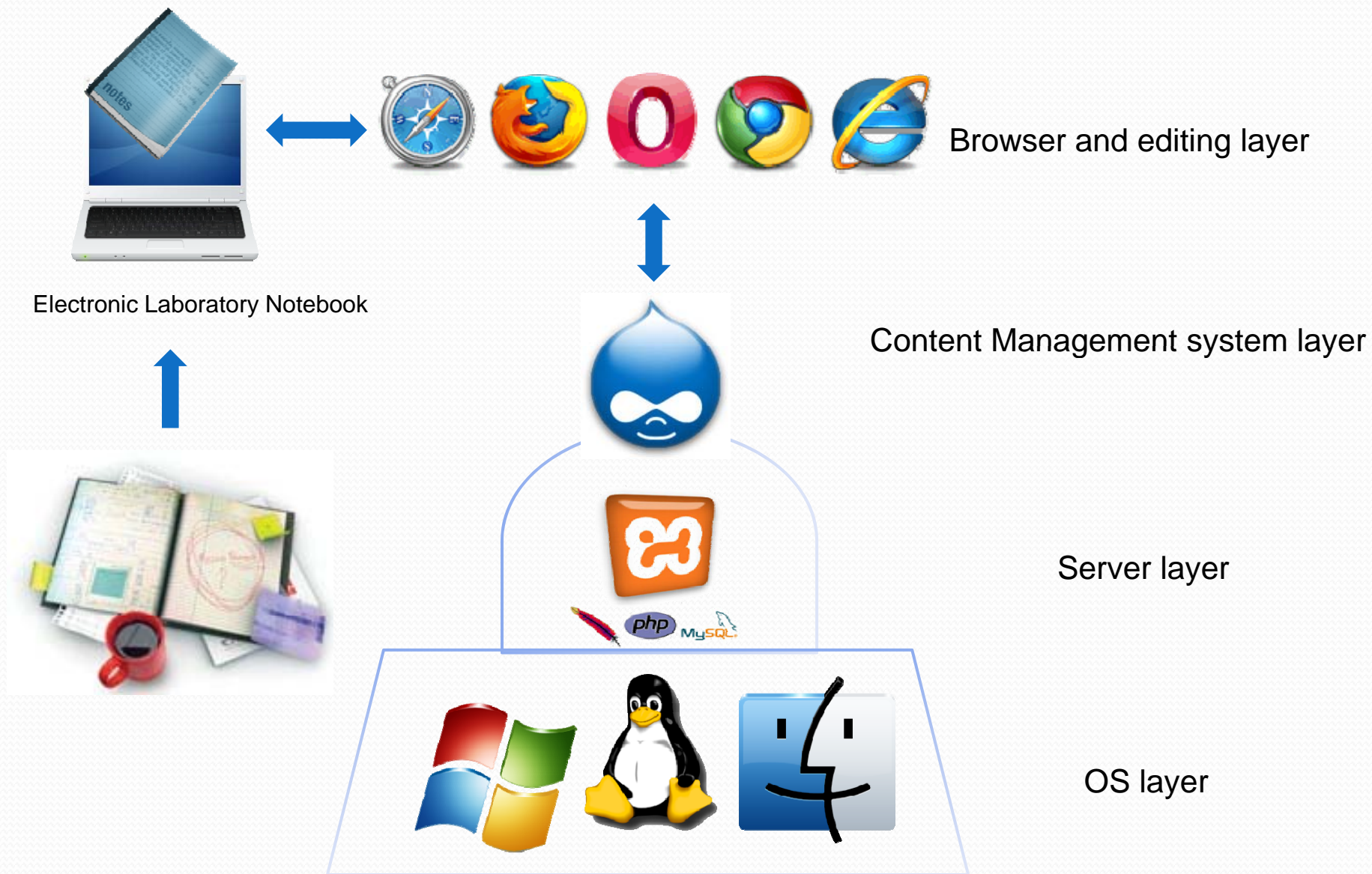


- Portable version

- Can be installed in windows/Mac and portable device with dynamic IP
- For personal use / small group with smooth learning curve



Basic Structure of ELN



Essential Functions in ELN



Friendly
Installation/ non
installation



Content generator



PDF printout



User management



Search



Web Access worldwide



Calendar/ Event



Webpage clip



Web share



Succinct control
panel



Image gallery



Backup /restore



Personalization



Digital signature



Security



Print



Data Exchange

Typical Interface for ELN

The screenshot shows a web interface for an ELN system. The header includes navigation links: 圖片瀏覽器, 搜尋, 行事曆, 討論區, 網路硬碟, 專案管理, 登出. The main title is 系統生物學暨網路生物學實驗室. The interface is divided into a left sidebar and a main content area.

Search: A search bar with a 搜尋 button is located in the top left of the sidebar.

Calendar/Coming event: A calendar for October 2008 is shown in the sidebar. The dates 1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 21, 22, 23, 24, 25, 26 are highlighted in yellow.

Taq: A list of tags is located at the bottom of the sidebar, including metagenomics lab, meeting ELN, and Bugs.

Record: The main content area displays a record titled "Poster for UPS, Hubba, Slides for PALM and ELN" with a MOLAS logo. The record details include: Start: 10/20/2008 - 00:00, End: 10/20/2008 - 23:59, Timezone: Etc/GMT, and a list of items: 1. UPS Poster, 2. Hubba Poster, 3. PALM Slides, 4. Eln Slides. A blue arrow points from the "Record" label to the MOLAS logo.

Response: Below the first record is another record titled "Electronic Notepad for iMac, iPhone" with a MOLAS logo. The record details include: Start: 10/02/2008 - 00:00, End: 10/02/2008 - 23:59, Timezone: Etc/GMT, and a URL: http://www.evernote.com/. A blue arrow points from the "Response" label to the MOLAS logo.

Additional Content: The sidebar contains sections for "建立內容" (Share this article, Latest articles, Post events, Post discussion articles, Upload archives, Create projects), "管理選單" (Modify site icon/info, Modify site title, User list), and "標籤集" (metagenomics lab, meeting ELN, Bugs).

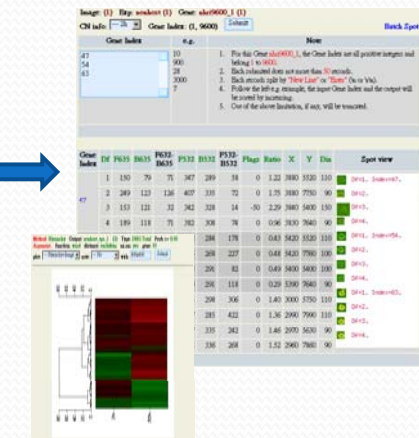
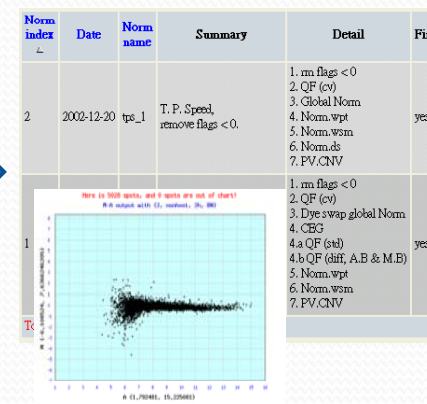
MOLAS



- MicroArray On Line Analysis System (MOLAS): a web-based customizable bioinformatics package designed for manager and analyze massive array data
- Remodeling to fit the need of transcriptomics generated from next generation sequencer (454, Selexa, SoLiD)

New Experiment Design

Upload raw data



Web interface

Exp - Experiment Data

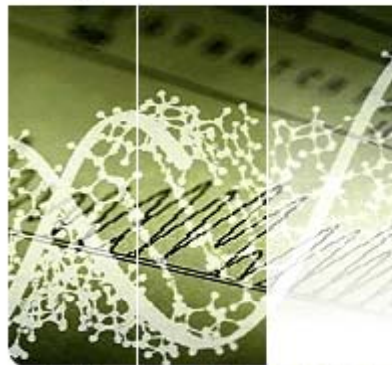
Data Normalization

Experiment Result and Analysis

Bioinformatics Core for Genomic Medicine and Biotechnology Development



GMBD Bioinformatics Core



Unit 1

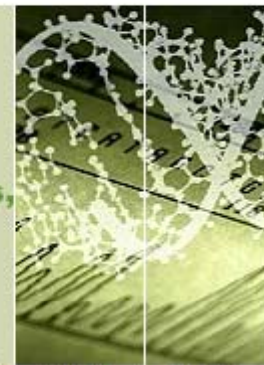
Unit 2

Unit 3

Comparative Genomics and Interactomes

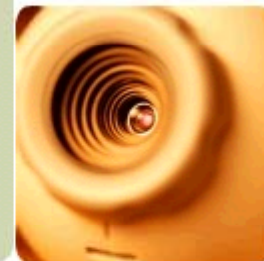
**Division of Bioinformatics and Biostatistics,
National Health Research Institutes**

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.



Unit 4

Unit 5



<http://www.tbi.org.tw>

Since 2002 Taiwan Bioinformatics Institute, All Rights Reserved.

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*



*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



系統生物與網路生物學實驗室
Laboratory of Systems Biology and Network Biology