
Curso de la Escuela Complutense de Verano 2005

Bioinformática y Biología Computacional

Redes de Interacciones entre Proteínas

Florencio Pazos

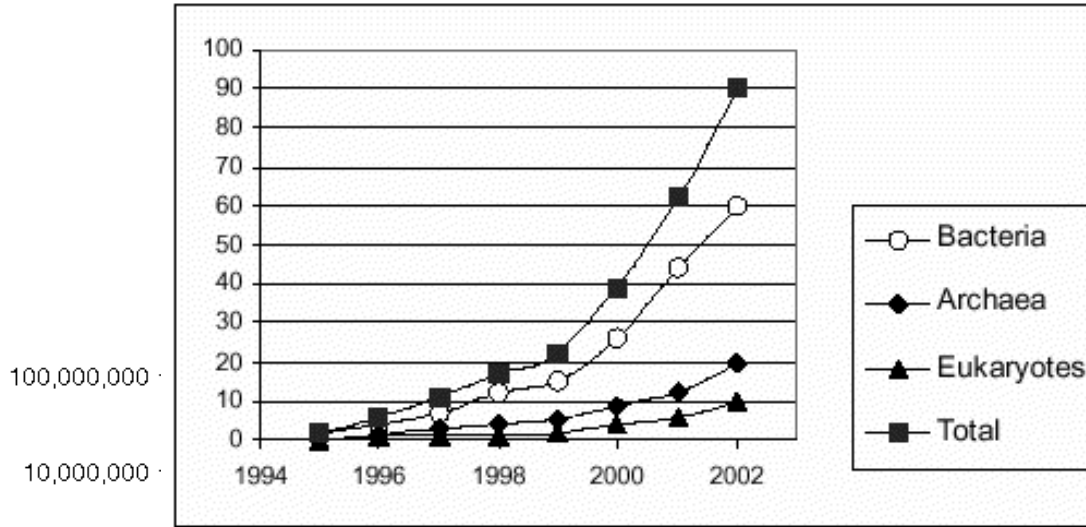
http://pdg.cnb.uam.es/pazos/cursos/Verano_UCM_05/

Florencio Pazos Cabaleiro
Protein Design Group (CNB-CSIC)
pazos@cnb.uam.es

Redes de Interacciones entre Proteínas

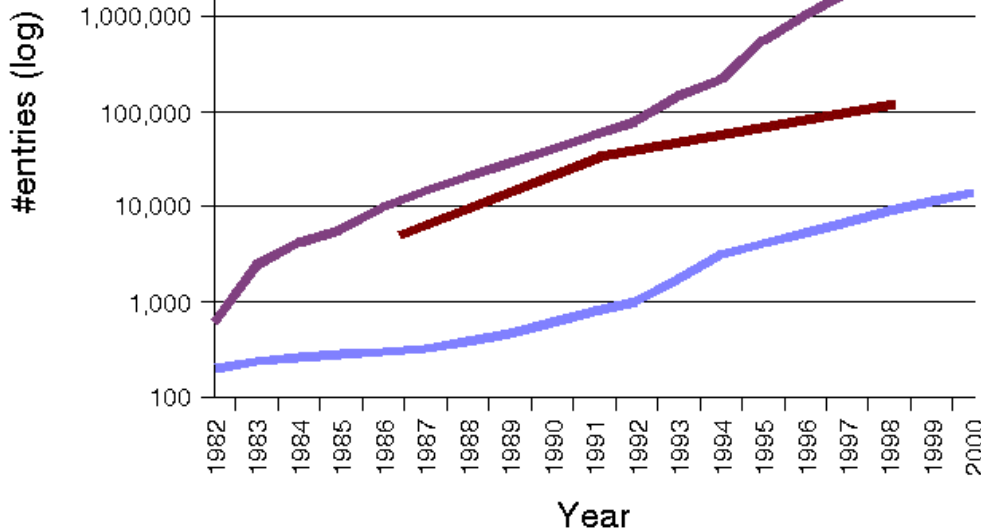
- Biología de sistemas y redes biológicas
- Determinación experimental a gran escala del interactoma
- Extracción automática de interacciones descritas en la literatura
- Predicción computacional de interacciones
- Propiedades de las redes de interacción
- Combinación con otra información
- *Bibliografía general*
- *Práctica*

Bioinformática orientada al componente individual



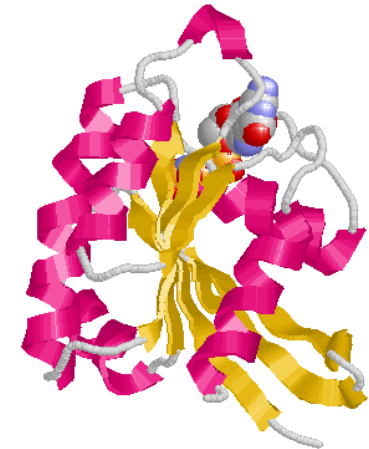
MREYKLVVLGSGGVGKSALTVQFVQGIFVDEYDPTIEDSY
 RKQVEVDCQQCMLEILDTAGTEQFTAMRDLYMKNGQGFAL
 VYSITAQSTFNDLQDLREQILRVKDTEDVPMILVGNKCDL
 EDERVVGKEQGQNLARQWCNCAFLESSAKSKINVNEIFYD
 LVRQINR

MLEILDTAGTEQFTAMRDLYMKNGQGFAL
 VYSITAQSTFNDLQDLREQILRVKDTEDVPMIL
 VGNKCDLEDERV



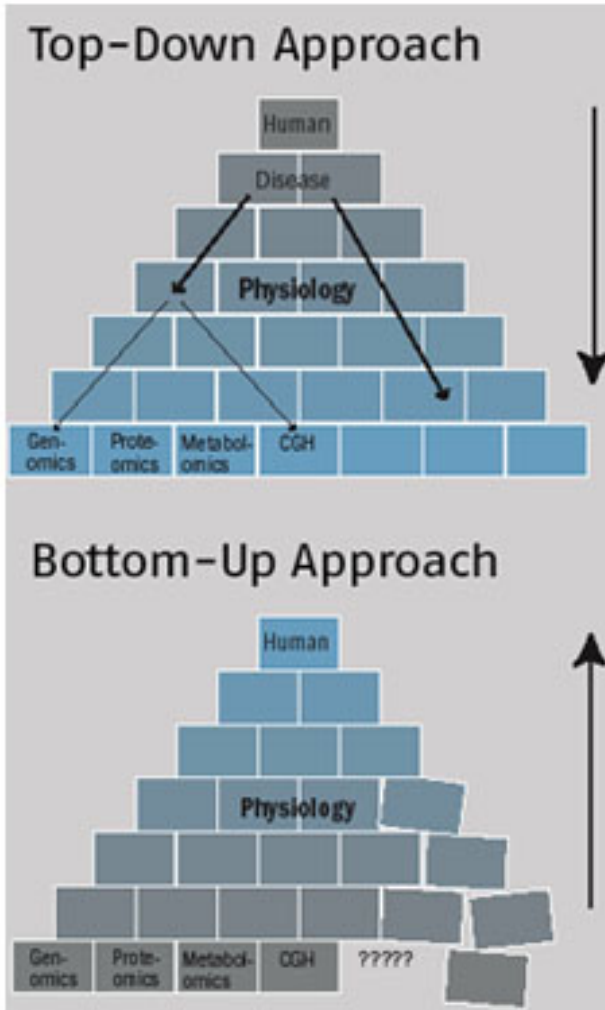
swissprot

PDB



Bioinformática y Sistemas Complejos en Biología

Biología de Sistemas vs. Biología Molecular



- Visión desde el punto de vista de sistemas complejos
- Objeto de estudio: **redes, relaciones, propiedades emergentes** (no propiedades de componentes individuales).
- Biología de Sistemas
- **Propiedades “globales” (emergentes, etc.)** -> conocimiento biológico
- Enfermedades (dianas/marcadores = **redes, patrones complejos**).



- Visión reduccionista
- Objeto de estudio: **componentes (genes proteínas)**
- **Biología Molecular**
- **Propiedades de los componentes** -> conocimiento biológico
- Enfermedades (dianas/marcadores = **proteínas/genes**).

¿Es suficiente la visión reduccionista?

- Reduccionismo en Biología muy exitoso (Biología Molecular). “*The ultimate aim of the modern movement in biology is to explain all biology in terms of physics and chemistry*”. F. Crick (1966)
- Sistemas biológicos: prototipo de sistemas complejos. => Muchos fenómenos biológicos nunca podrán explicarse a partir de las propiedades de los componentes (“el todo es mas que la suma de las partes”).
- Determinación de “repertorios de componentes” y sus características (secuenciación de genomas, proteómica, genómica estructural ...): Ni el número ni las características de genes y proteínas dan cuenta de muchas características de los organismos:
 - Similar número de genes en *Drosophila* y *C. elegans*.
 - Alta similaridad de secuencia entre humano y ratón.
 - ...

¿Es suficiente la visión reduccionista?

- Fallo en tratamiento de cáncer, ... En parte debido al enfoque reduccionista extremo.
- No vacuna HIV, ... “
- Fracaso vacunas de péptidos. “
- Reducción de fármacos en el mercado a pesar de la creciente inversión. (1 fármaco \leftrightarrow 1 diana).
- No resultados esperados para técnicas terapéuticas prometedoras de base reduccionista (terapia génica, RNA antisentido, ...).
- No mejora esperada de estas aproximaciones con secuenciación de genomas, etc.

¿Es suficiente la visión reduccionista?

Fallos en aproximaciones *in-vitro* e *in-silico*.

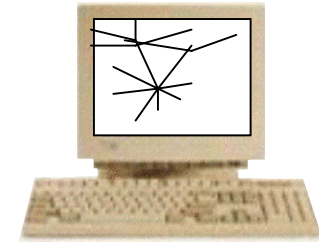
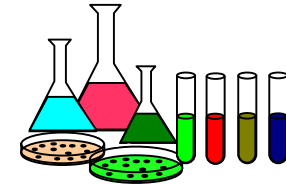
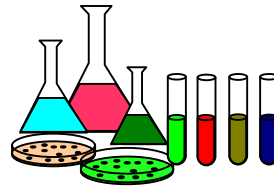
Fallos en técnicas experimentales de base reduccionista (delecciones, *knockout*, ...).

Knockout: no efecto, efecto distinto al esperado, o efecto “inespecífico” (cambio expresión 100's genes).

“Some mice should, by rights, be dead. At the very least, Teyumuras Kurzchalia expected his to be critically ill. But the most prominent symptom of his genetically engineered mice was a persistent erection”

Pearson, H. (2002) Surviving a knockout blow. *Nature*, **415**, 8-9.

Biología de componentes vs. Biología de sistemas



In vivo

Características del sistema

In vivo + in vitro

Características de los componentes (moléculas)

In vivo + in vitro + in silico

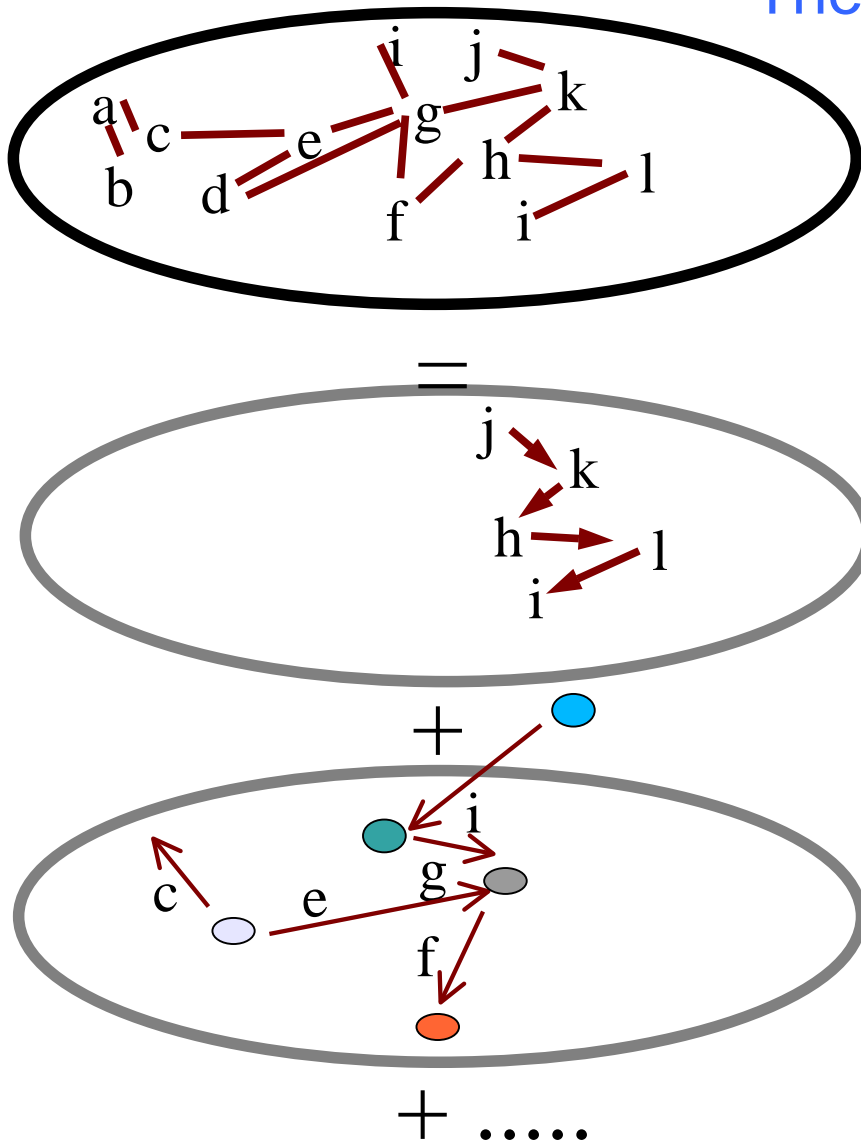
Características del sistema desde el pto. vista de los componentes y sus relaciones
Propiedades emergentes, ...

“genomics” / “post-genomics”

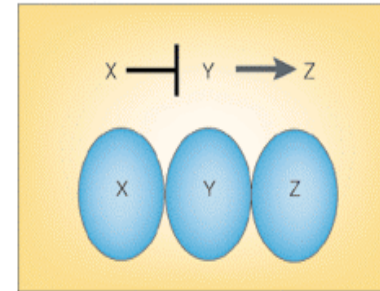
Multi-level high-throughput characterization of components

- Full-genome sequencing (“genome”).
- Characterization of transcripts (mRNA) (“transcriptome”)
- Characterization of the protein repertory (“proteome”)
- Cellular localization of the components (“localizome”)
- Genetic regulation networks (“regulome”)
- **Protein interaction networks (“interactome”)**
- High throughput characterization of gene-phenotype relationships (“phenome”)
- Metabolic networks (“metabolome”)
-

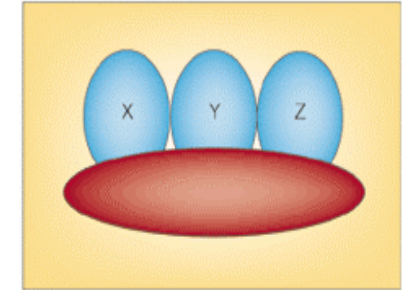
The interactome



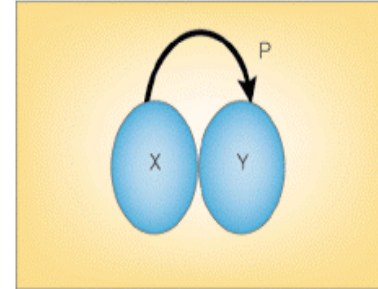
a Genetic pathways



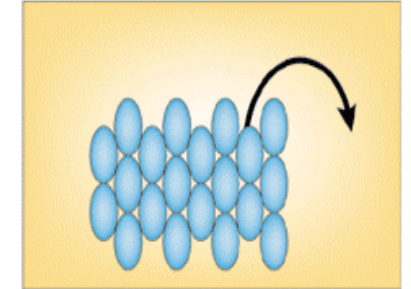
b Pathway scaffolding



c Enzymatic reactions

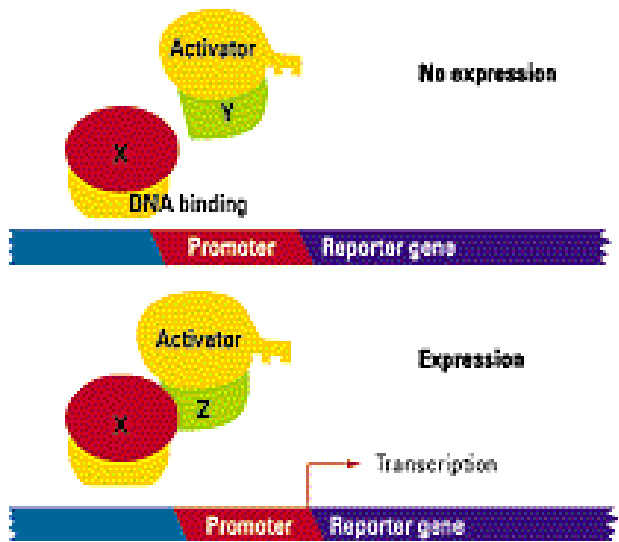


d Molecular machines

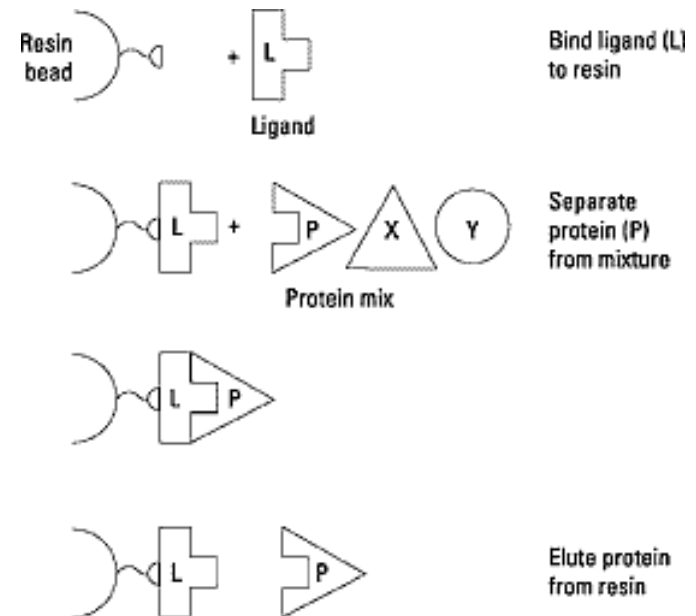


Nature Reviews | Molecular Cell Biology

Experimental determination of the interactome



Y2H

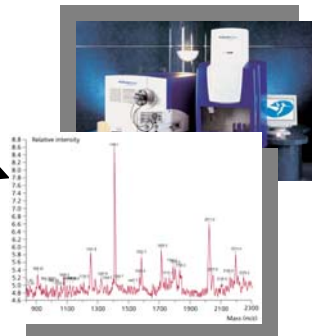
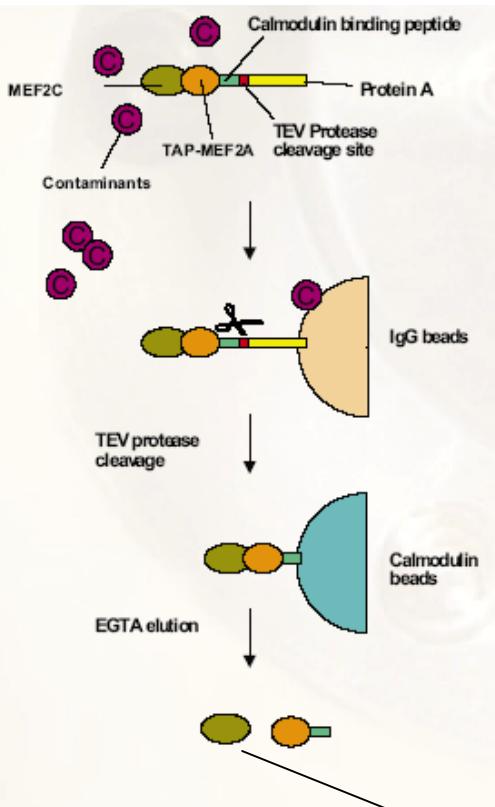


TAP/MS

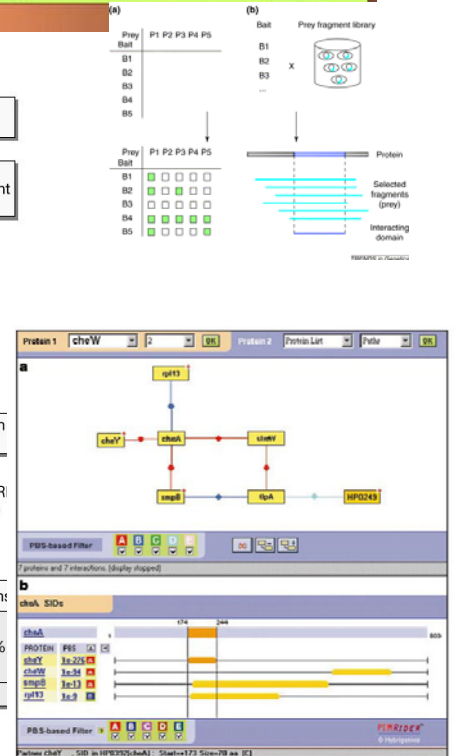
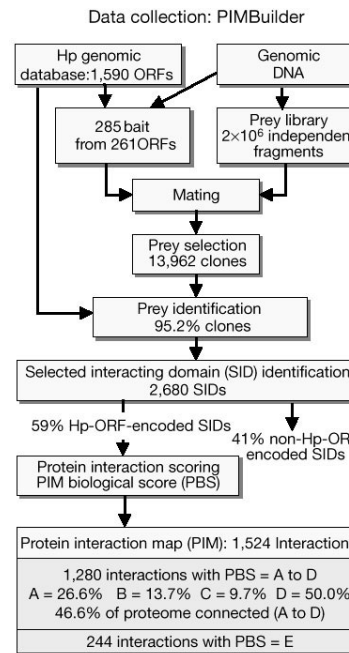
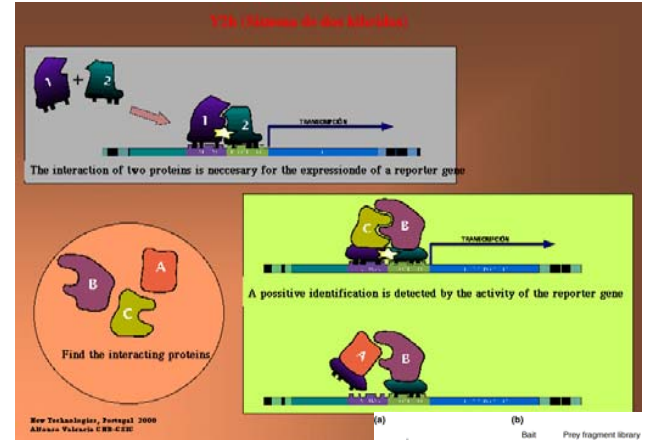
<http://pubs.acs.org/hotartcl/mdd/00/sep/edwards.html>

Experimental determination of the interactome

TAP/MS



Y2H



Quality of the high-throughput interaction data

Table 1
Large protein interaction screens for eukaryotes

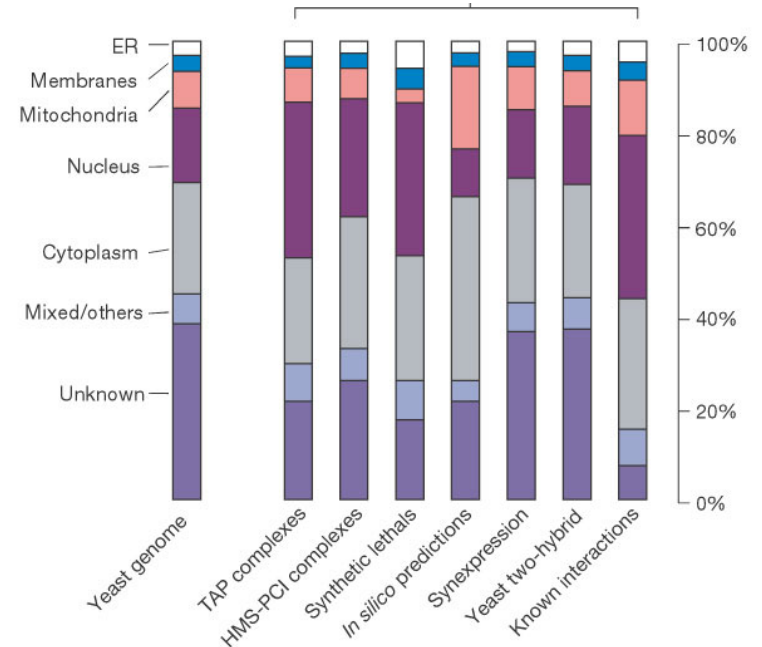
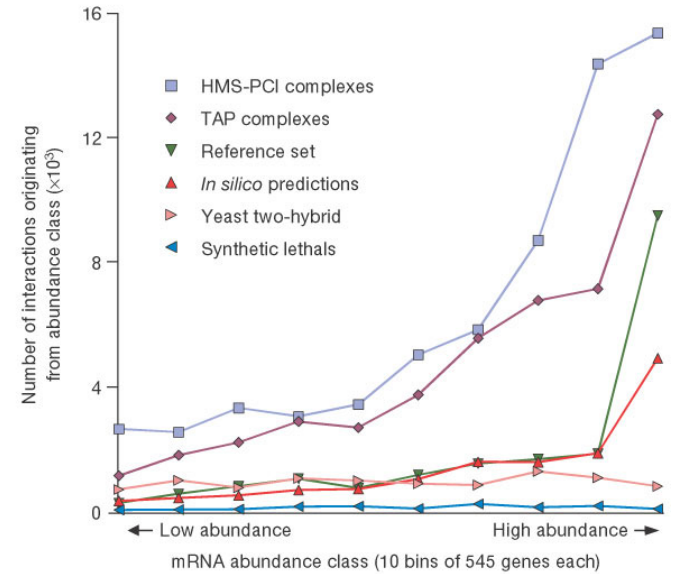
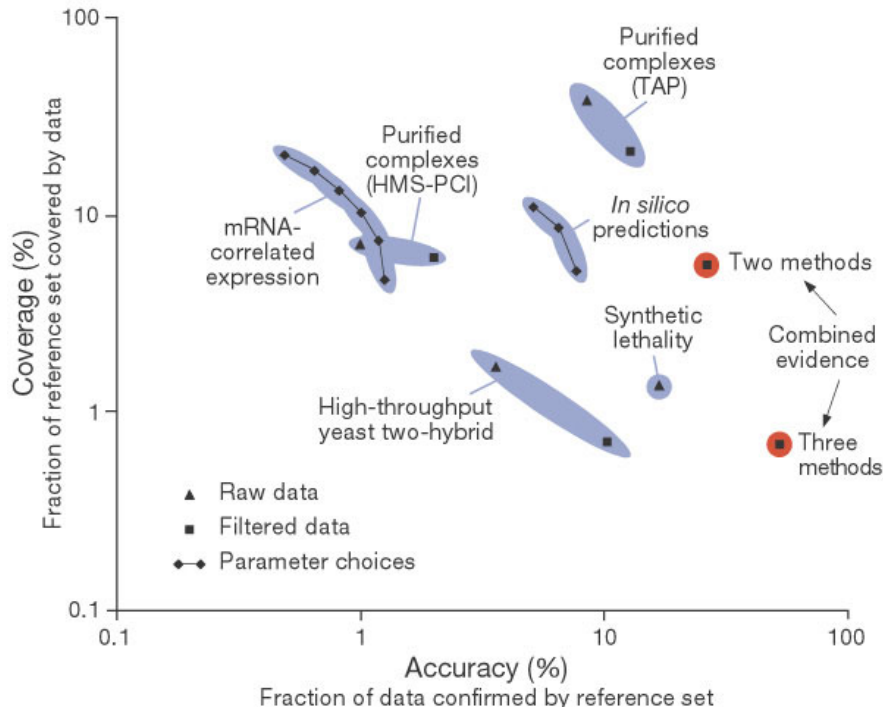
Organism (genes)	Method	Interactions ^a	Proteins	Reference
Yeast (~6000)	Yeast two-hybrid	967	1004	[63]
	Yeast two-hybrid	4549	3278	[13]
	Yeast two-hybrid	420	271	[64,65]
	Co-AP/MS	9421	1665	[66]
	Co-AP/MS	3878	1578	[67]
<i>Drosophila</i> (~14 000)	Yeast two-hybrid	20 405	7048	[49]
	Yeast two-hybrid	1814	488	[14]
Worm (~20 000)	Yeast two-hybrid	4027	1926	[68]

^aFor two-hybrid screens, the approximate number of unique binary interactions is shown. For co-AP/MS screens, the approximate number of binary interactions that would result if each bait protein contacted every protein that co-purified with it (the “hub and spoke” model) is shown. Data can be retrieved from one of the databases cited [42–44].

Overlap:
6 int !

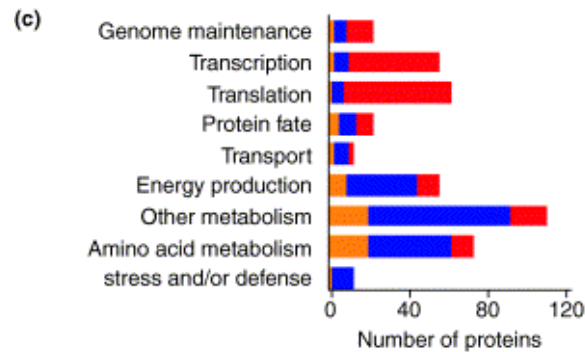
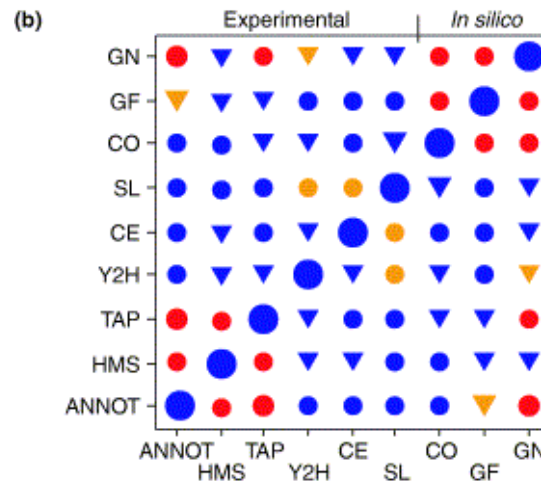
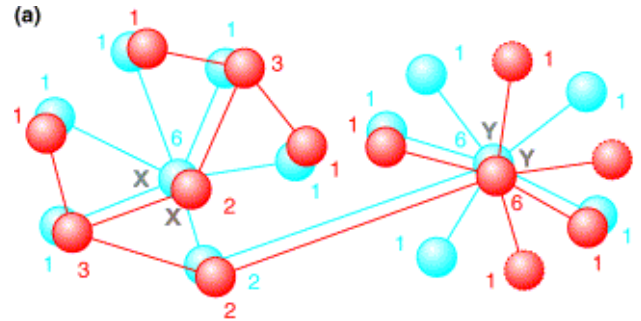
Estimation (yeast): 12.000-40000 (6000)

Quality of the high-throughput interaction data

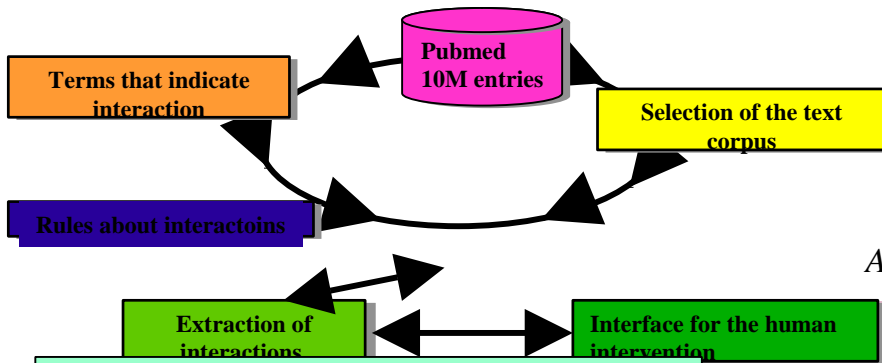


von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S. and Bork, P. (2002) Comparative assessment of large scale data sets of protein-protein interactions. *Nature*, **417**, 399-403.

Quality of the high-throughput interaction data



Retrieving protein relationships from the literature (text mining)

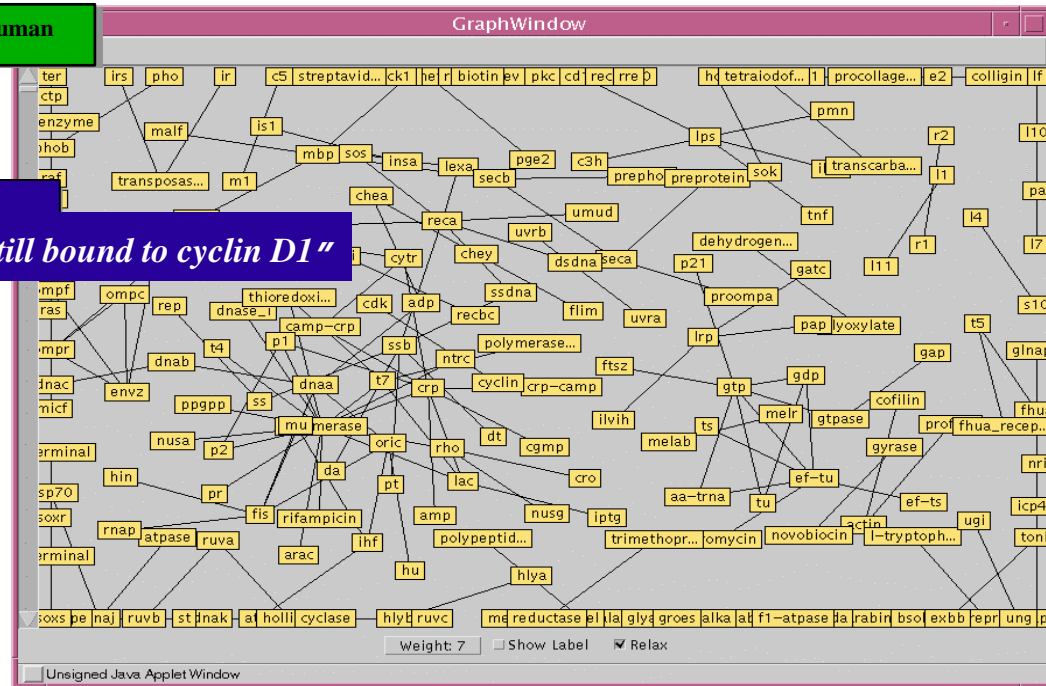


A. Valencia

Action words are for example:

~~bind, associate with, interact, phosphorylate, regulate~~

"After extensive purification, Cdk2 was still bound to cyclin D1"



Retrieving protein relationships from the literature. *iHop*

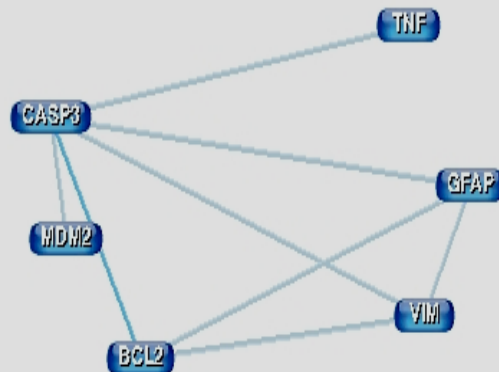
iHOP
 information hyperlinked
 Over Proteins

Search Gene

Clear model

Print version

Edges in the graph
 correspond to genes
 that concur in at least
 1 2 3 phrases.



MDM2 is **cleaved** by Caspase 3 (**CASP3**) during apoptosis after aspartic acid-361, generating a 60 kd fragment.

These findings indicate that IR-induced apoptosis involves activation of **CASP3** and that this CrmA-insensitive apoptotic pathway is distinct from those **induced** by **TNF** and certain other stimuli.

At the end of the experiment, the apoptotic markers (BAX, **BCL2** and

Fas-induced activation of the ce

iHOP
 information hyperlinked
 Over proteins

Search Gene

Show associations of
 SNF1 with genes
 from...

Human

Mouse

Drosophila

Zebrafish

C. elegans

Arabidopsis

S. Cerevisiae

E. Coli

Filter and options

Gene Model

Print version

Help

Symbol	Name	Synonyms	Organism
SNF1		CAT1, CCR1, GLC2, HAF3, PAS14	Saccharomyces cerevisiae

NCBI Protein NP_010765

The **snf1** mutation also **suppresses** the glucose repression defects of **reg1**.

The **SIP1** protein **co-immunoprecipitated** with **SNF1** and was phosphorylated in vitro.

Here we show that **Reg1** **interacts** with the **Snf1** catalytic domain in the two-hybrid system.

Previous studies showed that **Reg1** **regulates** the **Snf1** protein kinase in response to glucose.

The **SNF4** protein is physically **associated** with **SNF1** and positively affects the kinase activity.

The **Sip1** protein is known to undergo phosphorylation when **associated** in vitro with the **Snf1** protein kinase.

Genetic evidence indicated that the catalytic activity of **Snf1** negatively **regulates** its interaction with **Reg1**.

The **SNF1** protein kinase and the **associated SNF4** protein are required for release of glucose repression in Saccharomyces cerevisiae.

The **SIP1** gene of Saccharomyces cerevisiae is a carbon-catabolite-specific negative regulator of GAL gene transcription and acts as a multicopy suppressor of growth defects **associated** with impaired **Snf1p** protein kinase activity.

We show that different sequences of **Reg1** **interact** with **Glc7** and **Snf1**.

In two-hybrid assays, one **SNF4** mutation **enhances** the interaction between **Snf4** and **Snf1**.

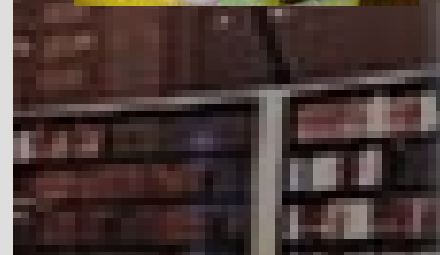
Previously, we identified **SIP1** and **SIP2** as proteins that **interact** with **SNF1** in vivo by the two-hybrid system.

Previous experimental evidence had indicated that **Reg1** might **target** **Glc7** to nuclear substrates such as the **Snf1** kinase complex.

The catalytic subunits of Arabidopsis SnRKs, AKIN10 and AKIN11, interact with **Snf4** and **suppress** the **snf1** and **snf4** mutations in yeast.

Pak1 **associates** with the **Snf1** kinase in vivo, and the association is greatly enhanced under glucose-limiting conditions when **Snf1** is active.

We show that **SNF4** **binds** to the **SNF1** regulatory domain in low glucose, whereas in high glucose the regulatory domain binds to the kinase domain of **SNF1** itself.



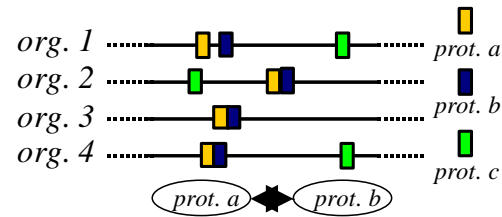
In-silico prediction of protein interactions

a) phylogenetic profiles

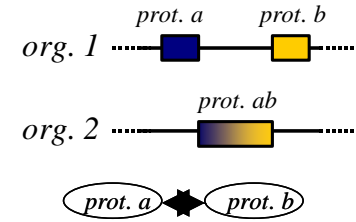
	prot. a	prot. b	prot. c	prot. d
org. 1	1	1	1	1
org. 2	0	1	0	1
org. 3	1	0	1	0
org. 4	1	0	1	1

(prot. a) ↔ (prot. c)

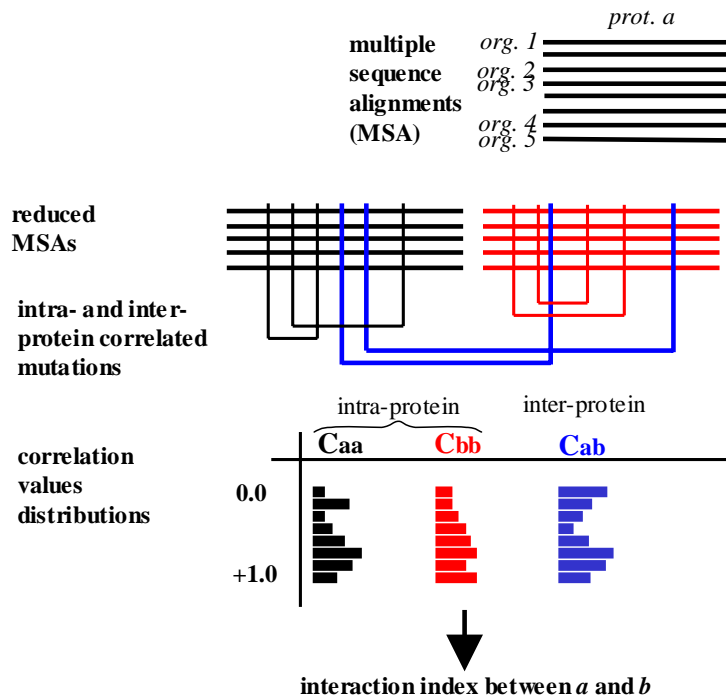
b) conservation of gene neighbouring



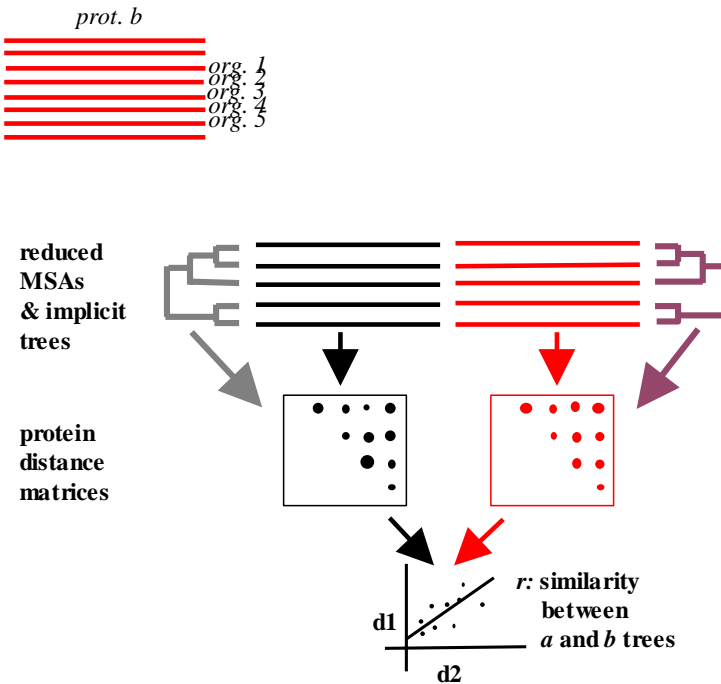
c) gene fusion



e) correlated mutations

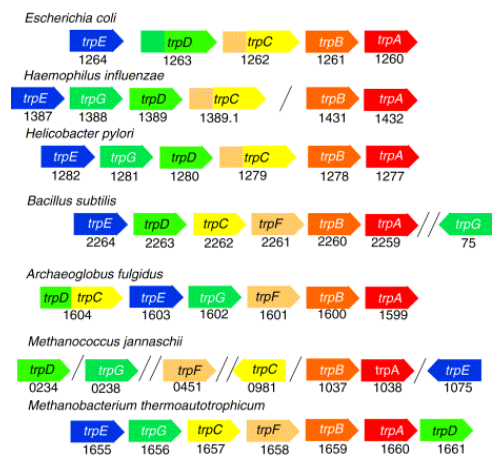
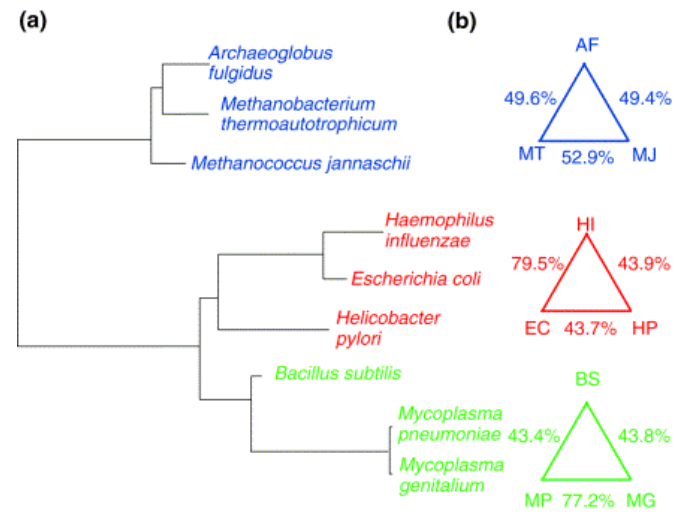
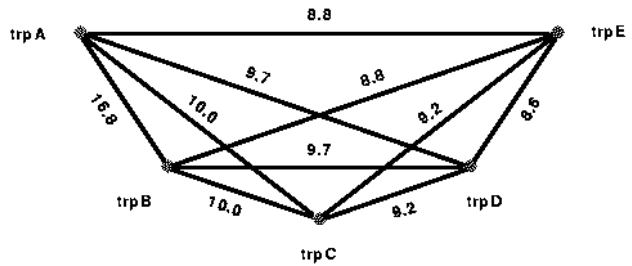
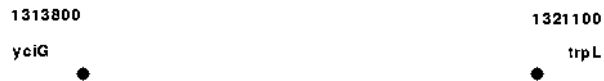
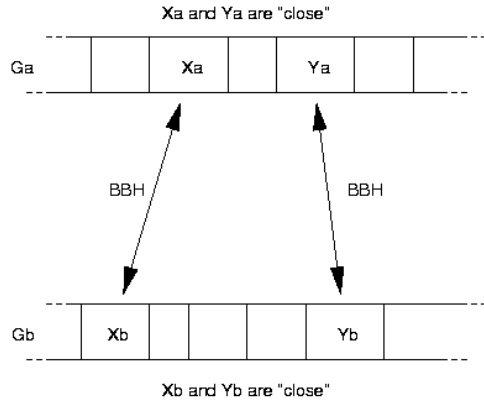


d) similarity of phylogenetic trees



- Huynen, M., Snel, B., Lathe, W. & Bork, P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res*, **10**, 1204-1210.
- Valencia, A. & Pazos, F. (2002) Computational methods for the prediction of protein interactions. *Curr Opin Struct Biol*, **12**, 368-373.
- Salwinski, L. & Eisenberg, D. (2003). Computational methods of analysis of protein-protein interactions. *Curr Opin Struct Biol*. **13**, 377-382.

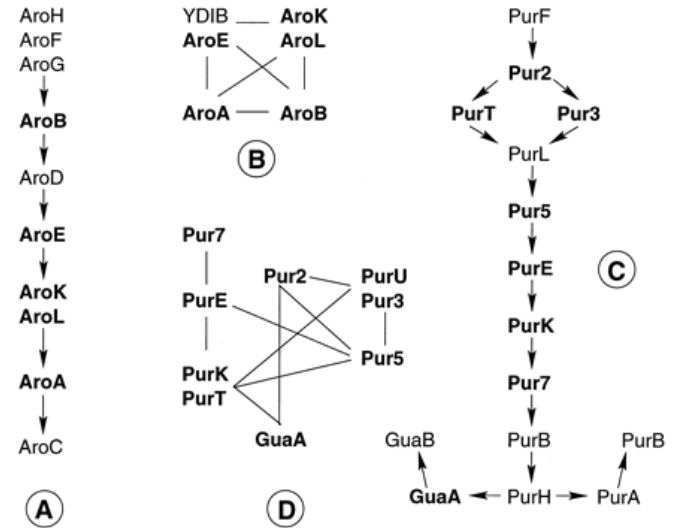
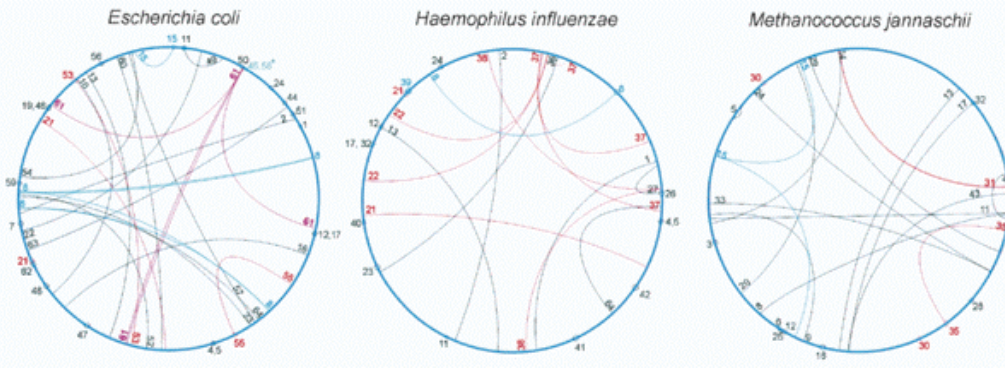
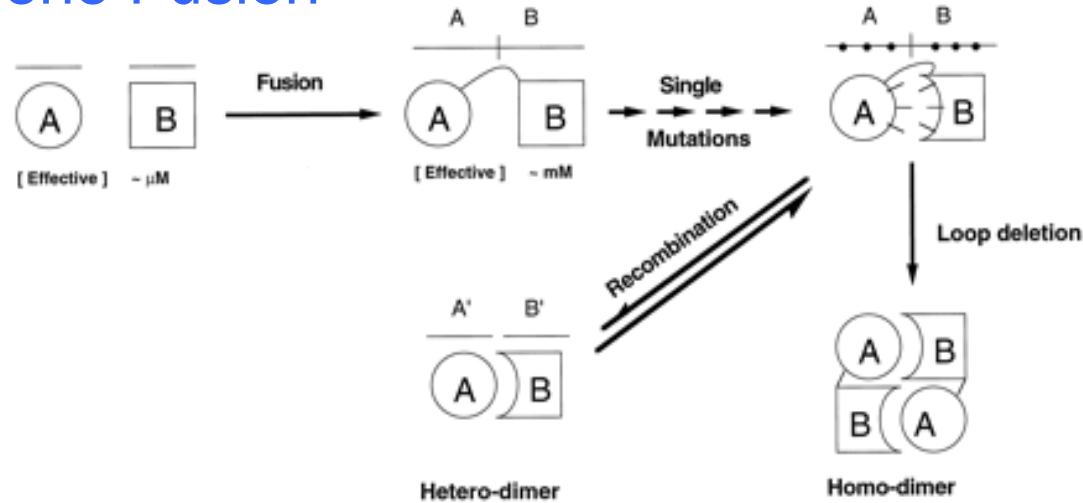
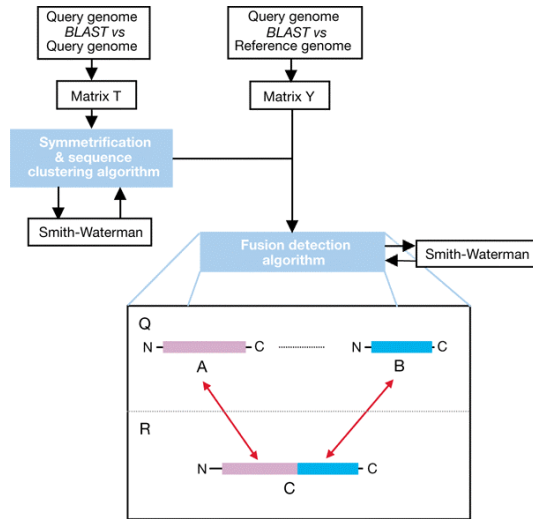
Conservation of gene neighboring



Overbeek, R., Fonstein, M., D'Souza, M., Pusch, G. D. & Maltsev, N. (1999). Use of contiguity on the chromosome to predict functional coupling. *In Silico Biol.* **1**, 93-108.

Dandekar, T., Snel, B., Huynen, M. & Bork, P. (1998). Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci.* **23**, 324-328.

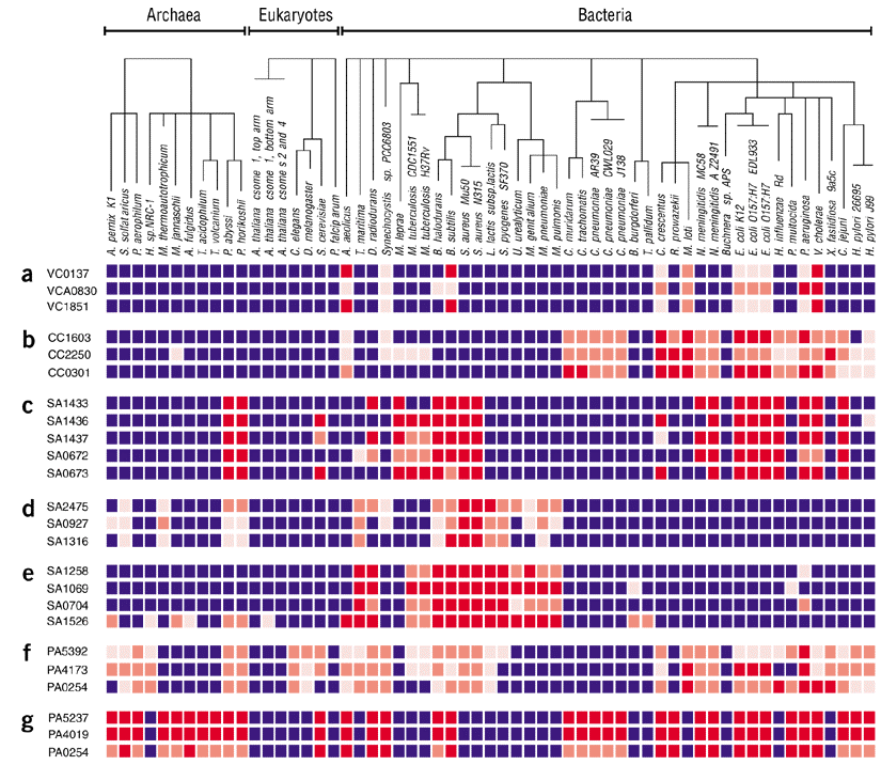
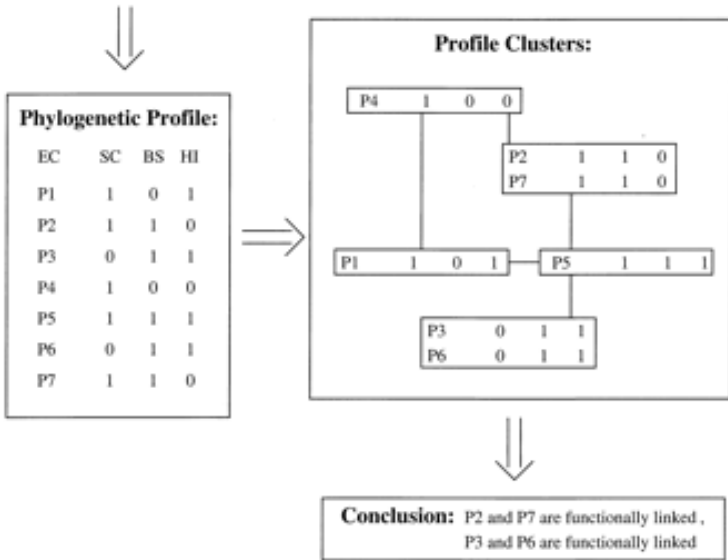
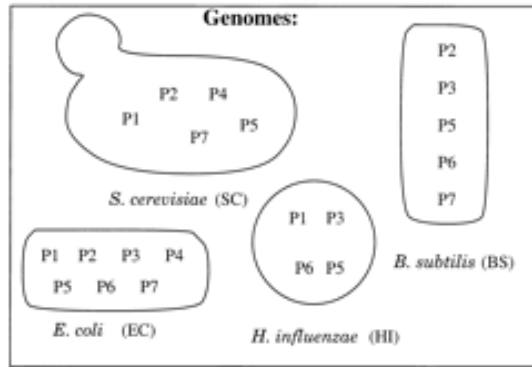
Gene Fusion



Enright, A. J., Iliopoulos, I., Kyripides, N. C. & Ouzounis, C. A. (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*. **402**, 86-90.

Marcotte, E. M., Pellegrini, M., Ho-Leung, N., Rice, D. W., Yeates, T. O. & Eisenberg, D. (1999). Detecting protein function and protein-protein interactions from genome sequences. *Science*. **285**, 751-753.

Phylogenetic Profiles



$$p_{ij} = -1/\log E_{ij}$$

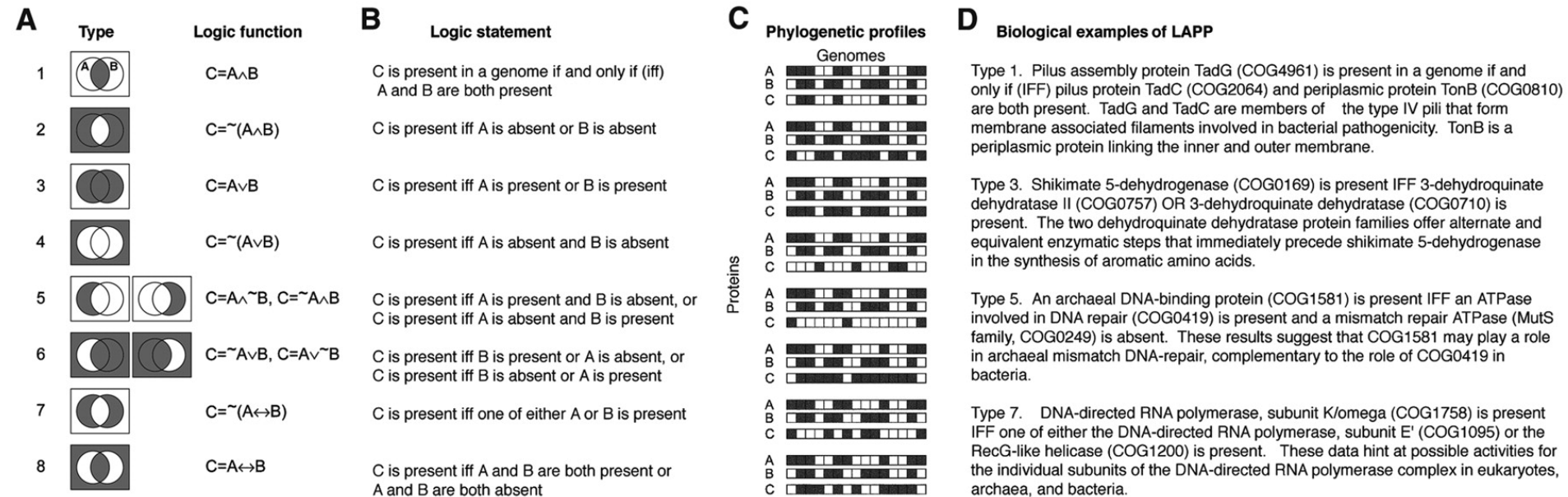
$$H(A) = - \sum p(a) \ln p(a)$$

$$MI(A,B) = H(A) + H(B) - H(A,B)$$

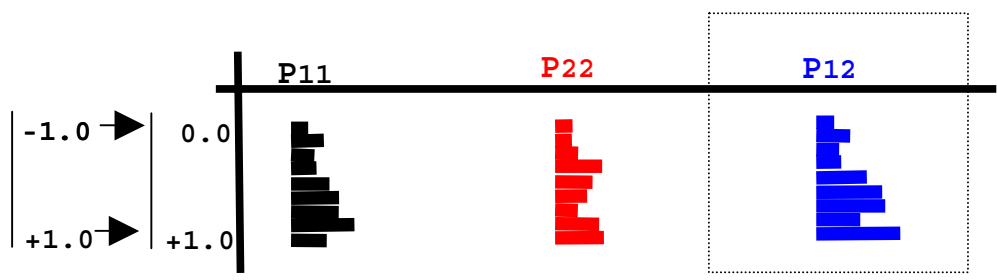
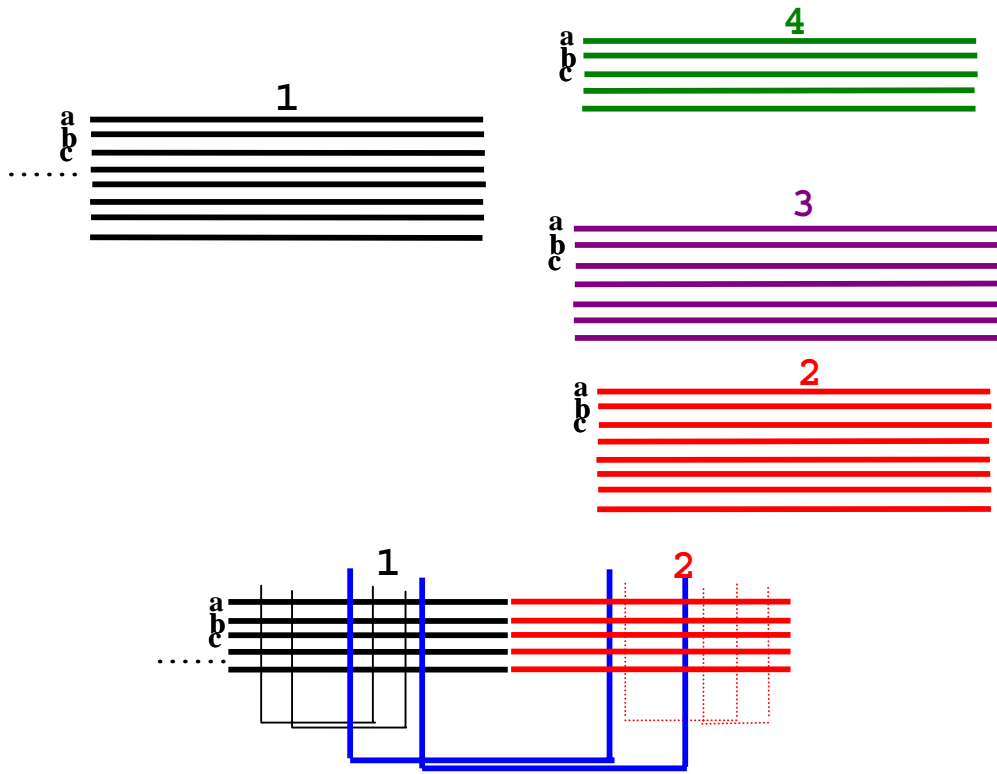
•Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. (1999). Assigning protein functions by comparative genome analysis: Protein pylogenetic profiles. *Proc Natl Acad Sci USA*. **96**, 4285-4288.

•Date, S. V. & Marcotte, E. M. (2003). Discovery of uncharacterized cellular systems by genome-wide analysis of functional linkages. *Nat Biotechnol*. **21**, 1055-1062.

Phylogenetic Profiles



Based on correlated mutations (*i2h*)

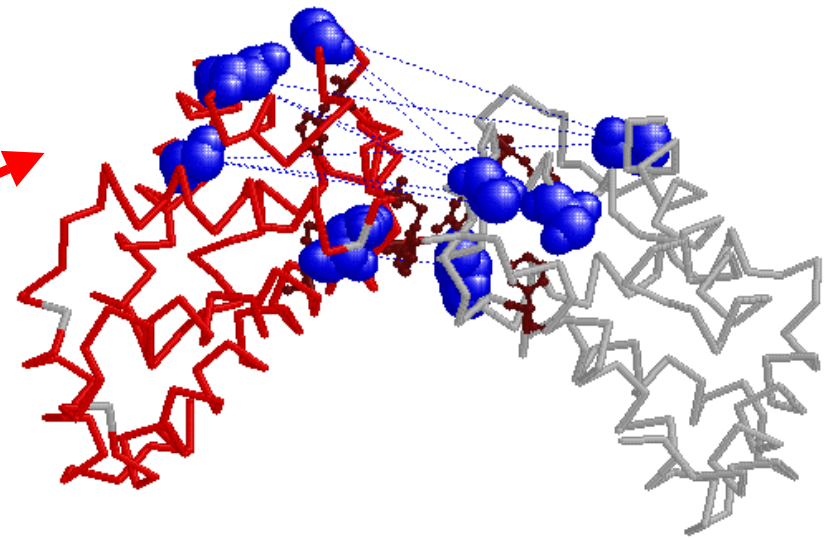
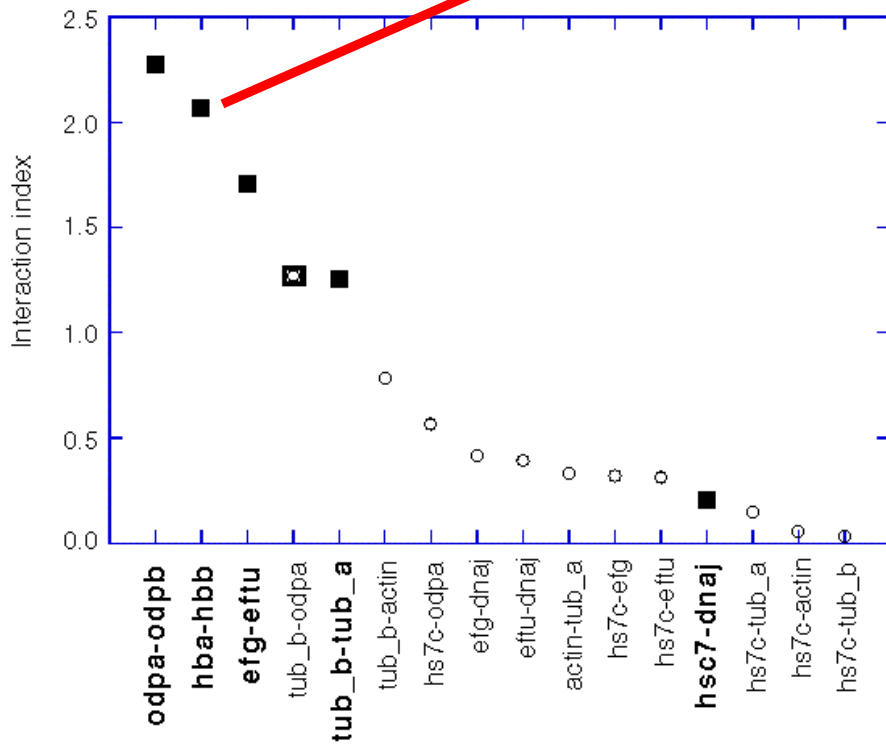


$$C_{12} = \sum_{i=incorr}^{1.0} \frac{P_{12i}}{P_{11i} + P_{22i}} \cdot i$$

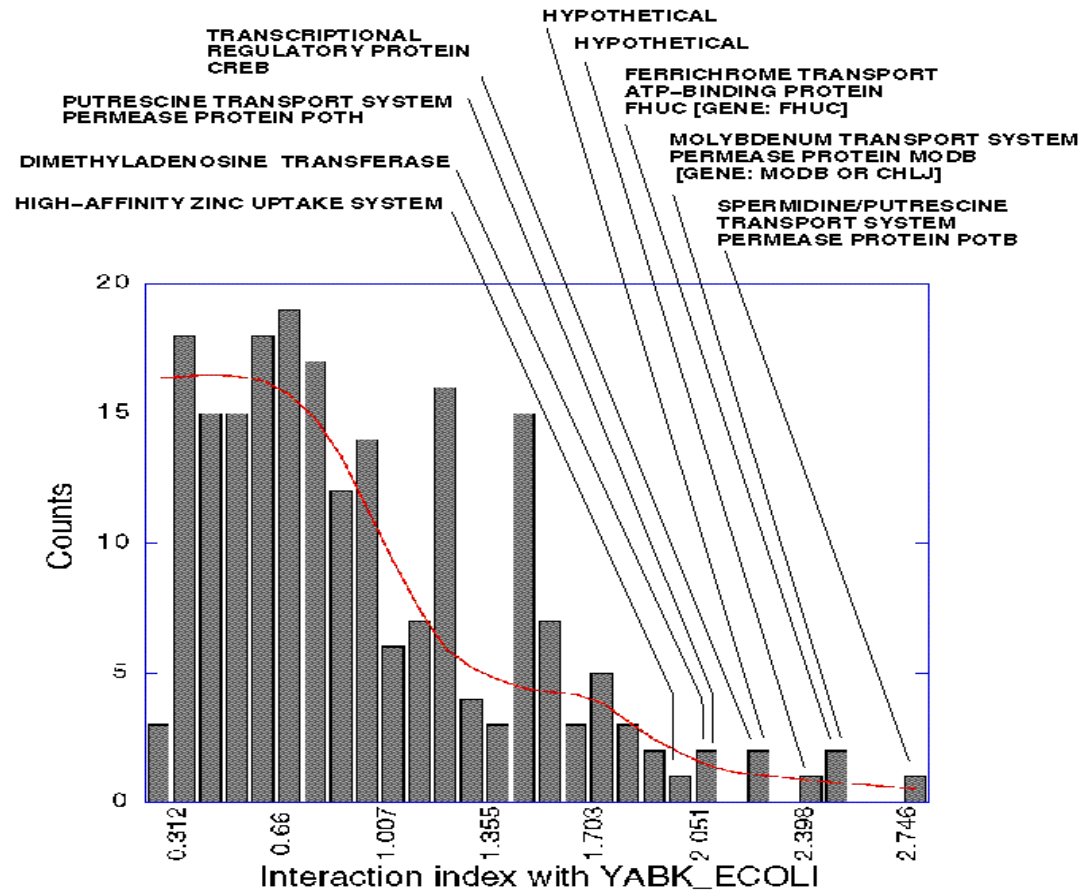
Based on correlated mutations (i2h)

2c2c_2-1alc_1		3,503	1alc_1-1rnd_1		0,968
1sgt_2-4mt2_1		3,448	3adk_2-4tnc_2		0,961
9pap_1-9pap_2	*	3,042	1alc_1-1rnd_2		0,957
1alc_1-1alc_2	*	2,852	1sgt_1-2c2c_2		0,889
2c2c_1-4mt2_1		2,825	2c2c_2-3pgk_1		0,878
4tms_1-4tms_2	*	2,735	3trx_1-9pap_2		0,857
3trx_1-3trx_2	*	2,571	4tnc_1-4mt2_2		0,853
4mt2_1-4mt2_2	*	2,469	4tnc_2-4mt2_2		0,836
2c2c_2-4mt2_1		2,355	3trx_1-3pgk_2		0,829
2c2c_2-4mt2_2		2,331	3trx_1-9pap_1		0,814
4tnc_1-4tnc_2	*	2,238	2c2c_2-1rnd_2		0,813
3blm_1-3blm_2	*	2,206	4tms_2-3dfr_2		0,809
3pgk_1-3pgk_2	*	2,197	9pap_2-3adk_2		0,805
2c2c_1-4mt2_2		2,139	4tms_1-3dfr_2		0,804
1sgt_2-2c2c_1		2,068	1sgt_2-1alc_1		0,799
2c2c_1-1alc_1		2,011	9pap_1-3adk_2		0,790
2c2c_1-1alc_2		1,886	3trx_2-9pap_2		0,761
3adk_1-3adk_2	*	1,862	4tnc_2-4mt2_1		0,747
1sgt_2-2c2c_2		1,835	3adk_2-3pgk_2		0,726
2c2c_1-2c2c_2	*	1,787	4tnc_1-4mt2_1		0,718
3adk_1-3pgk_1		1,624	9pap_2-4tnc_2		0,702
1rnd_1-4mt2_1		1,530	3trx_1-3adk_1		0,673
2c2c_1-9pap_2		1,520	3dfr_1-3dfr_2	*	0,657
3adk_2-3dfr_2		1,507	2pf2_2-1alc_2		0,628
1sgt_2-2pf2_2		1,489	3adk_1-4tnc_1		0,617
9pap_1-3adk_1		1,488	3adk_1-4tnc_2		0,614
3adk_1-3pgk_2		1,444	2pf2_2-1alc_1		0,595
2c2c_2-1alc_2		1,415	3adk_2-4tnc_1		0,539
2c2c_1-3pgk_2		1,389	4tms_1-3dfr_1		0,507
1sgt_1-4mt2_1		1,387	3trx_2-3pgk_1		0,489
3adk_1-3dfr_1		1,367	3trx_2-3pgk_2		0,471
1rnd_2-4mt2_1		1,359	3trx_1-3adk_2		0,471
2c2c_2-3adk_1		1,319	1sgt_1-1alc_1		0,455
1rnd_1-1rnd_2	*	1,314	3trx_1-2c2c_2		0,453
3pgk_1-4tms_1		1,299	3trx_1-2c2c_1		0,446
2c2c_1-3adk_1		1,297	4tms_2-4tnc_2		0,444
3pgk_1-4tms_2		1,292	2c2c_1-1rnd_2		0,442
3trx_1-3pgk_1		1,279	1sgt_2-1alc_2		0,435
2c2c_1-3pgk_1		1,278	3trx_2-3adk_1		0,427
1alc_1-4mt2_1		1,278	4tms_1-4tnc_2		0,413
2c2c_2-9pap_2		1,274	1sgt_1-1rnd_1		0,403
1rnd_1-4mt2_2		1,258	4tms_1-4tnc_1		0,401
3adk_2-3pgk_1		1,252	4tms_2-3dfr_1		0,398
1rnd_2-4mt2_2		1,240	1alc_2-4mt2_2		0,362
3adk_1-3dfr_2		1,209	1sgt_1-1rnd_2		0,358
3trx_2-2c2c_1		1,196	1sgt_1-4mt2_2		0,356
3pgk_2-4tms_2		1,178	1sgt_2-1rnd_1		0,352
3pgk_2-4tms_1		1,170	3trx_1-4tnc_2		0,316
			2c2c_1-4tnc_1		0,303

Based on correlated mutations (*i2h*)

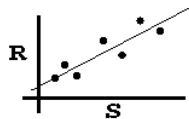
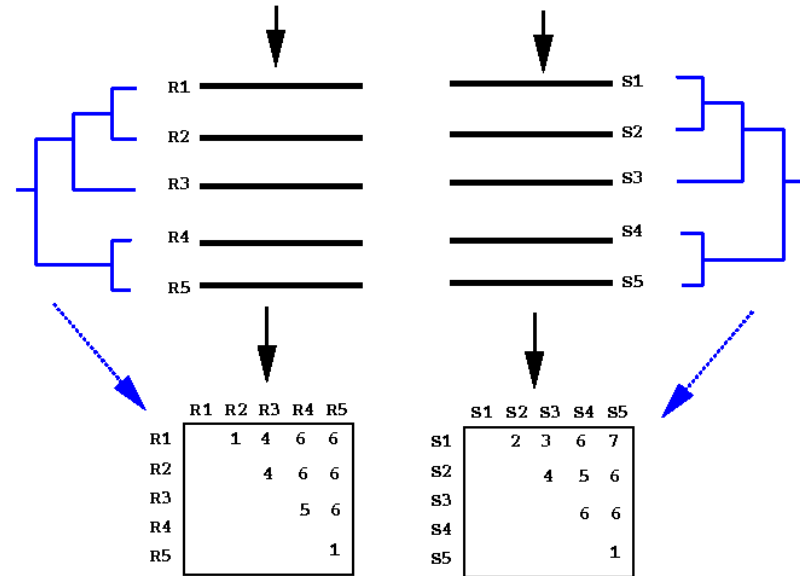
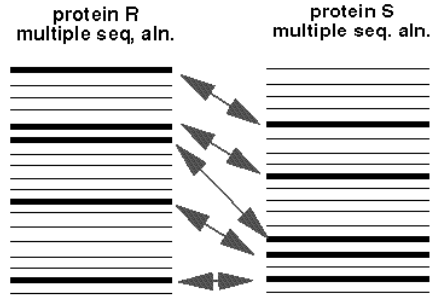


Interaction-based function prediction



- Pazos, F. & Valencia, A. (2002). In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins*. **47**, 219-227.
- Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. (2003). Global protein function prediction from protein-protein interaction networks. *Nat Biotechnol*. **21**, 697-700.
- Samanta, M. P. & Liang, S. (2003). Predicting protein functions from redundancies in large-scale protein interaction networks. *Proc Natl Acad Sci U S A*. **100**, 12579-12583.

MirrorTree

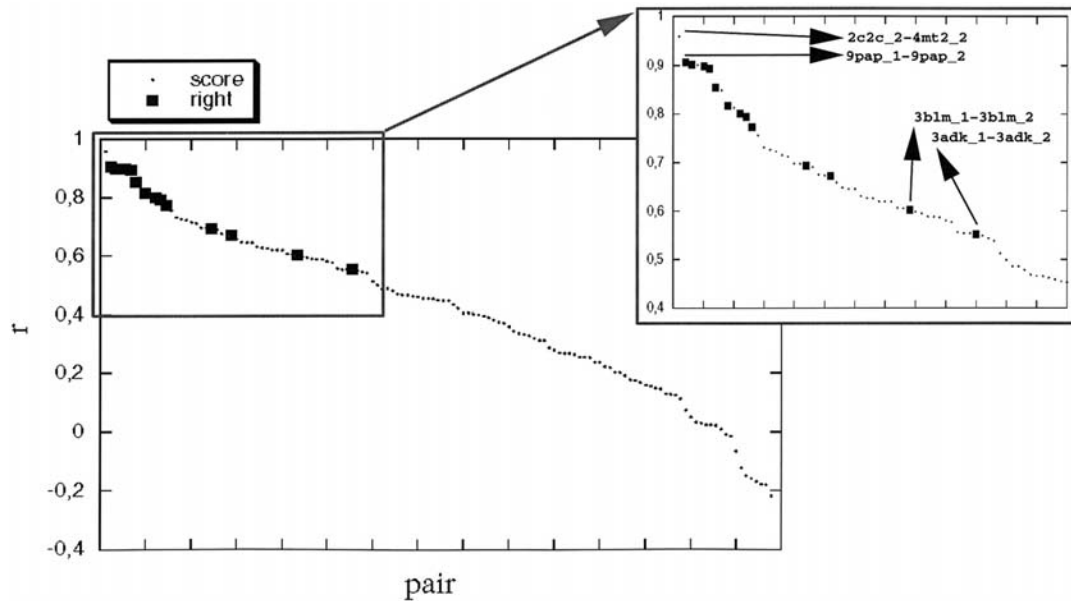


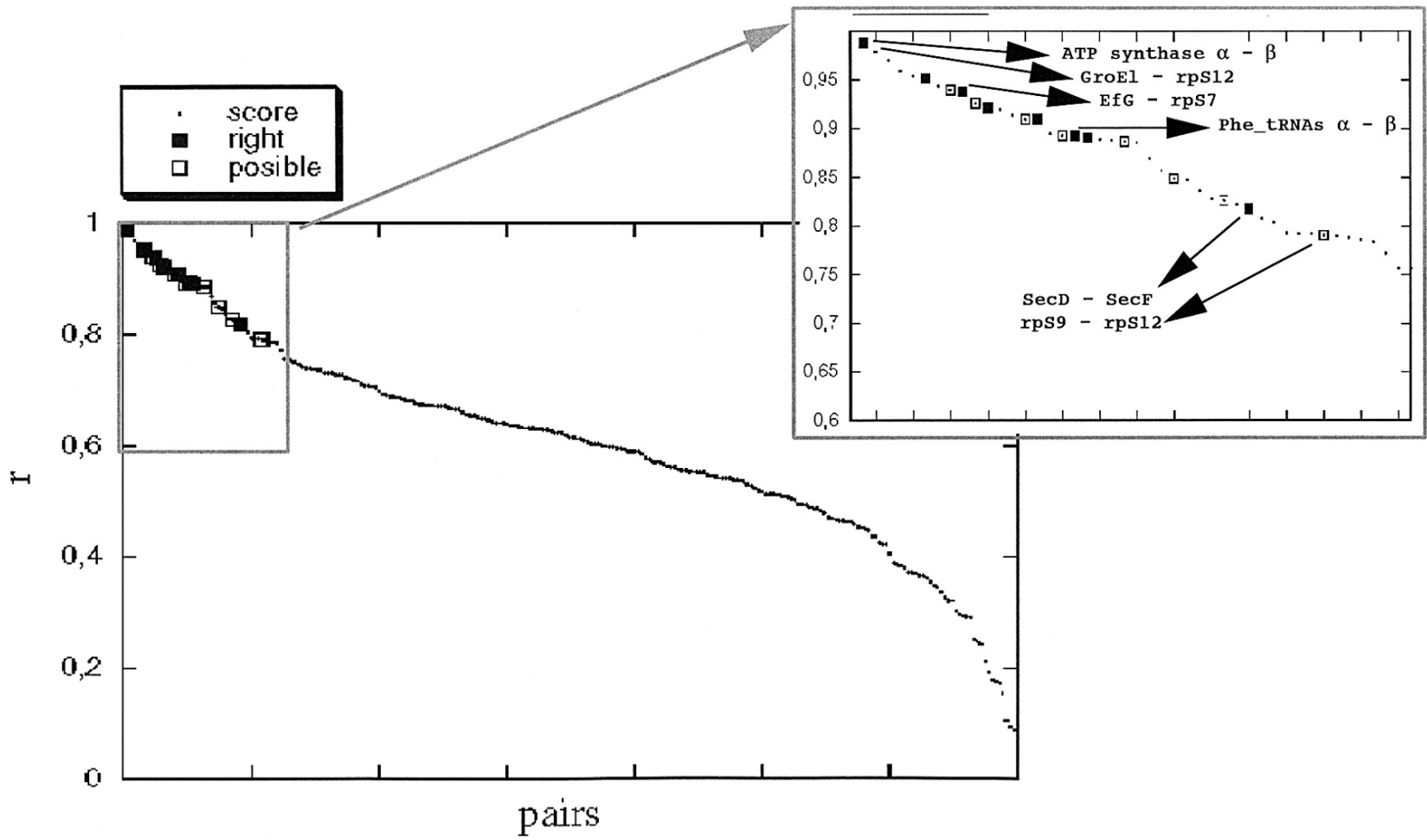
$$r = \frac{\sum_{i=1}^n (R_i - \bar{R}) \cdot (S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \cdot \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

Goh, C.-S., Bogan, A.A., Joachimiak, M., Walther, D. and Cohen, F.E. (2000) Co-evolution of Proteins with their Interaction Partners. *J Mol Biol*, **299**, 283-293.

Pazos, F. and Valencia, A. (2001) Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng*, **14**, 609-614.

2c2c_2-4mt2_2	0,959	4tnc_1-4mt2_2	0,448
9pap_1-9pap_2 *	0,907	9pap_2-4tnc_1	0,446
2c2c_1-4mt2_2	0,901	1alc_2-4mt2_2	0,446
3pgk_1-3pgk_2 *	0,901	1sgt_2-4mt2_1	0,433
4mt2_1-4mt2_2 *	0,898	3adk_1-4tnc_2	0,421
3trx_1-3trx_2 *	0,894	1rnd_1-4mt2_2	0,405
4tms_1-4tms_2 *	0,854	4tnc_2-4mt2_2	0,405
2c2c_2-4mt2_1	0,849	2c2c_1-3adk_2	0,401
1rnd_1-1rnd_2 *	0,817	1sgt_2-2c2c_1	0,399
2c2c_1-4mt2_1	0,813	4tms_2-3dfr_2	0,394
1alc_1-1alc_2 *	0,801	3adk_1-3dfr_1	0,390
4tnc_1-4tnc_2 *	0,794	1sgt_2-2c2c_2	0,381
2c2c_1-2c2c_2 *	0,773	3adk_2-3dfr_1	0,372
3pgk_1-4tms_1	0,756	1sgt_2-1alc_1	0,371
3pgk_1-4tms_2	0,731	4tms_1-3dfr_2	0,358
2c2c_1-3adk_1	0,726	1sgt_1-4mt2_1	0,343
3pgk_2-4tms_1	0,723	1sgt_1-4mt2_2	0,336
2c2c_2-3pgk_1	0,715	9pap_2-4tnc_2	0,331
1alc_1-1rnd_1	0,712	4tms_1-3dfr_1	0,327
2c2c_2-3pgk_2	0,698	3trx_1-2c2c_2	0,319
1alc_2-1rnd_1	0,697	3trx_1-2c2c_1	0,312
1sgt_1-1sgt_2 *	0,693	1sgt_1-1alc_1	0,312
3pgk_2-4tms_2	0,691	3trx_2-2c2c_2	0,287
3adk_2-3dfr_2	0,675	3trx_2-2c2c_1	0,281
1sgt_2-2pf2_2	0,673	1sgt_1-2c2c_2	0,270
3dfr_1-3dfr_2 *	0,672	1alc_1-4mt2_2	0,268
2c2c_2-9pap_1	0,658	1sgt_1-2c2c_1	0,268
2c2c_1-3pgk_1	0,648	2c2c_1-1rnd_1	0,263
3trx_2-9pap_1	0,646	9pap_1-3adk_2	0,254
1sgt_1-2pf2_2	0,646	2c2c_2-3adk_2	0,254
2c2c_2-3adk_1	0,631	3adk_2-3pgk_1	0,251
3trx_1-9pap_1	0,627	1sgt_1-1rnd_1	0,238
2c2c_2-1alc_2	0,626	3adk_2-3pgk_2	0,238
2c2c_1-3pgk_2	0,620	9pap_2-3adk_2	0,221
3trx_2-9pap_2	0,620	1sgt_2-1alc_2	0,219
1rnd_2-4mt2_1	0,619	2c2c_2-1alc_1	0,203
1alc_2-1rnd_2	0,607	9pap_1-4tnc_1	0,202
1rnd_2-4mt2_2	0,606	1sgt_2-1rnd_1	0,191
3blm_1-3blm_2 *	0,603	1sgt_1-1alc_2	0,178
1alc_1-1rnd_2	0,599	3trx_2-3adk_2	0,175
3trx_1-3pgk_1	0,595	1sgt_1-1rnd_2	0,168
3trx_1-9pap_2	0,589	2pf2_2-1alc_1	0,160
1alc_2-4mt2_1	0,588	2c2c_1-1alc_1	0,155
2c2c_1-1alc_2	0,587	9pap_1-4tnc_2	0,149
2c2c_1-9pap_1	0,581	2c2c_2-1rnd_2	0,146
3trx_1-3pgk_2	0,577	4tms_2-3dfr_1	0,130
4tnc_1-4mt2_1	0,556	3trx_1-3adk_2	0,128
3adk_1-3pgk_1	0,554	2c2c_2-1rnd_1	0,125
		2c2c_1-1rnd_2	0,113

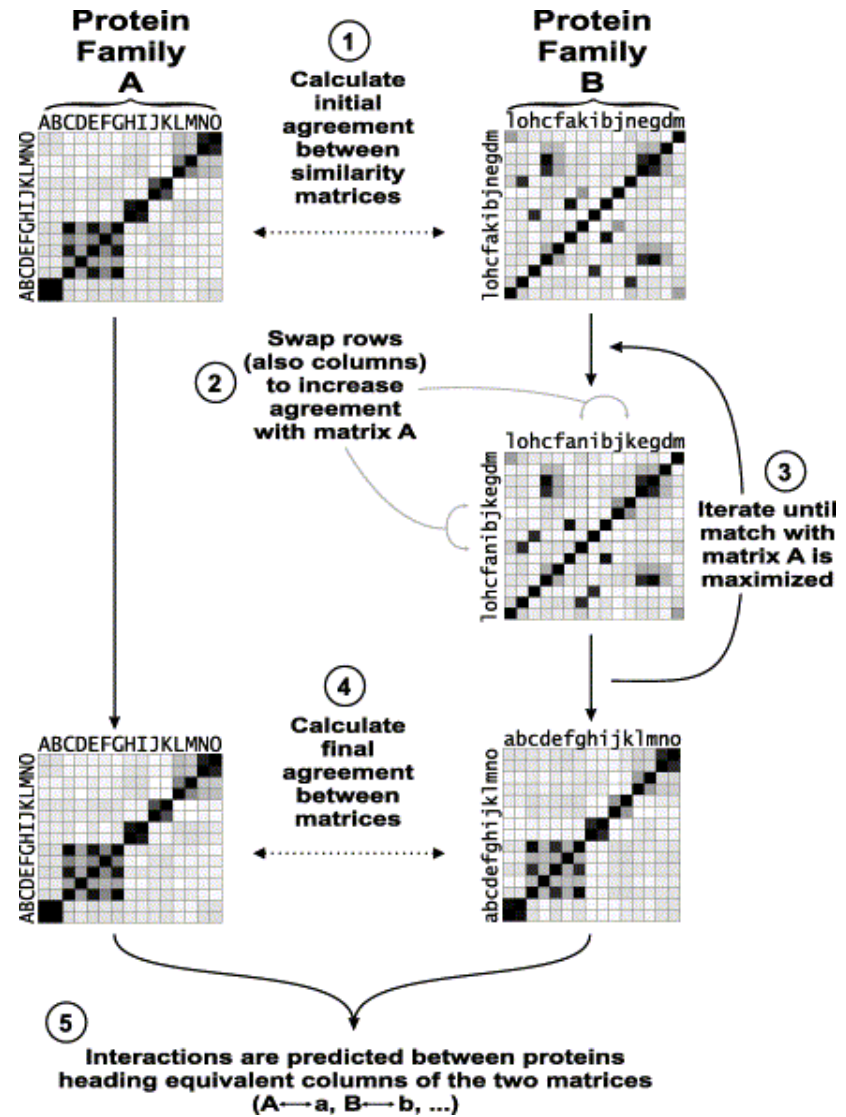
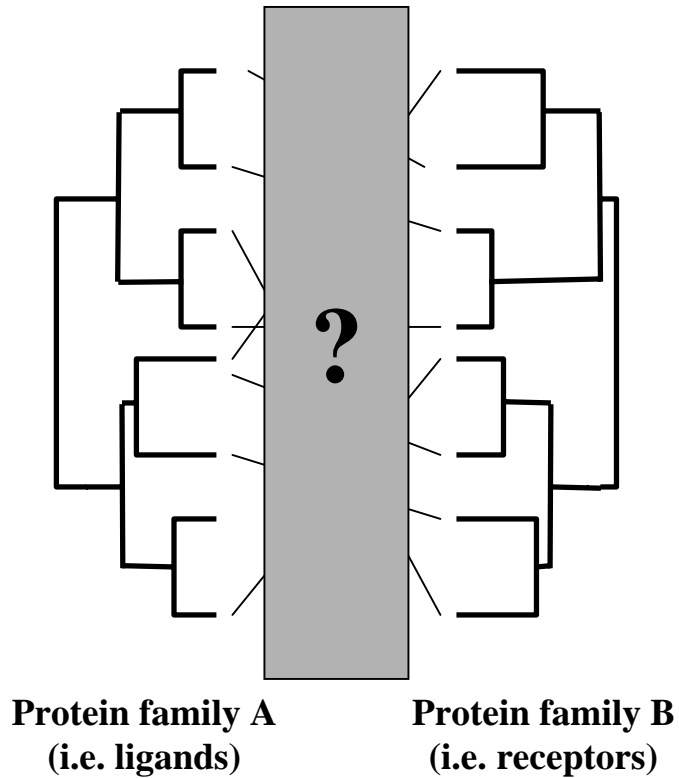




MirrorTree Variations

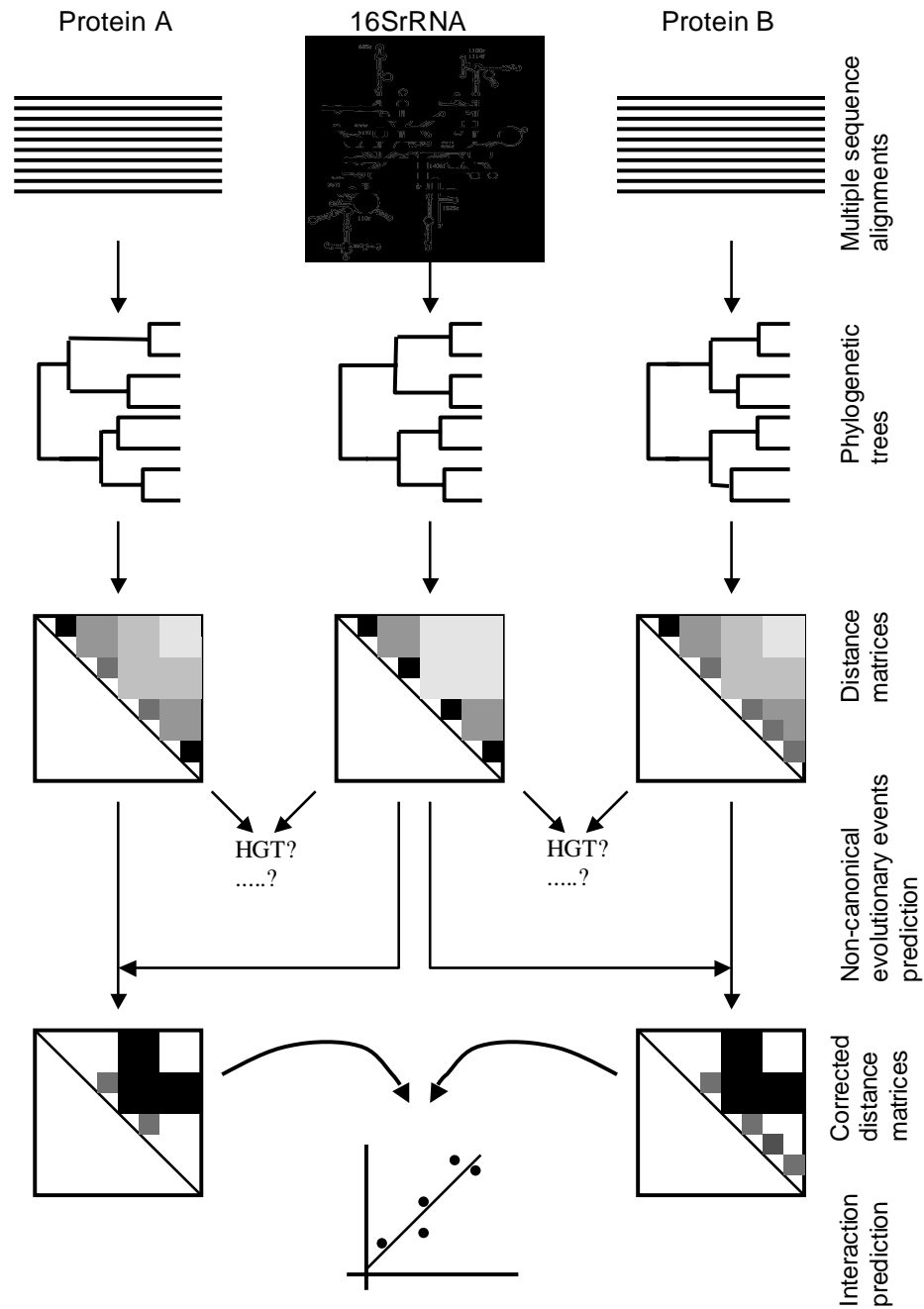
- Gertz, J., Elfond, G., Shustrova, A., Weisinger, M., Pellegrini, M., Cokus, S. and Rothschild, B. (2003) Inferring protein interactions from phylogenetic distance matrices. *Bioinformatics*, **19**, 2039-2045.
- Goh, C.S. and Cohen, F.E. (2002) Co-evolutionary analysis reveals insights into protein-protein interactions. *J Mol Biol*, **324**, 177-192.
- Ramani, A.K. and Marcotte, E.M. (2003) Exploiting the co-evolution of interacting proteins to discover interaction specificity. *J Mol Biol*, **327**, 273-284.
- Sato, T., Yamanishi, Y., Horimoto, K., Toh, H. and Kanehisa, M. (2003) Prediction of protein-protein interactions from phylogenetic trees using partial correlation coefficient. *Genome Informatics*, **14**, 496-497.
- Kim, W.K., Bolser, D.M. and Park, J.H. (2004) Large-scale co-evolution analysis of protein structural interlogues using the global protein structural interactome map (PSIMAP). *Bioinformatics*, **20**, 1138-1150. Epub 2004 Feb 1135.
- Tan, S., Zhang, Z. and Ng, S. (2004) ADVICE: Automated Detection and Validation of Interaction by Co-Evolution. *Nucl. Acids. Res.*, **32**, W69-W72.

MirrorTree. Variations



MirrorTree
Variations

tol-mirrortree



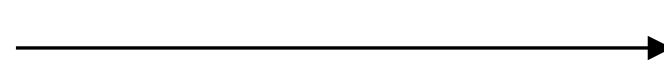
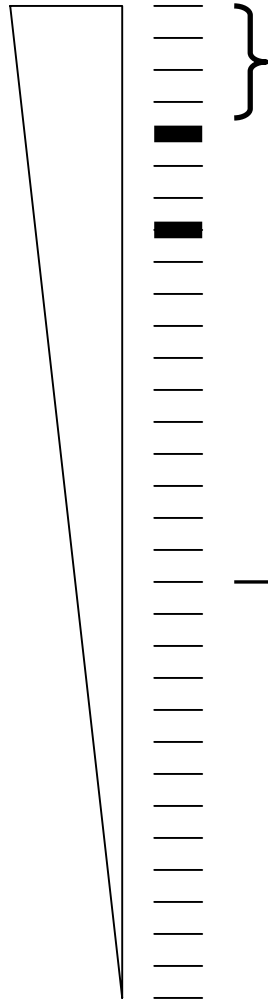
tol-mirrortree

DIP:
516 interactions (*E coli*)

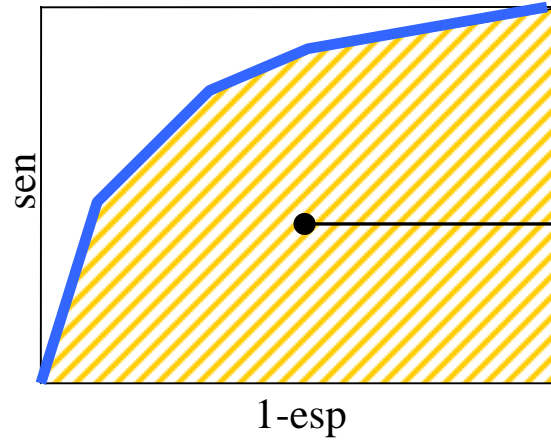
20,087 pairs calculated
(115 true)

118 proteins with
 ≥ 1 calculated true interactor

P00000
List of pairs
sorted by score

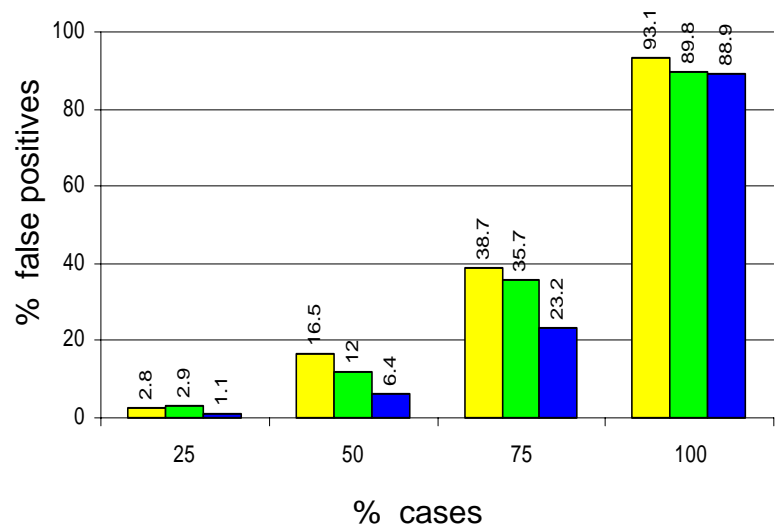
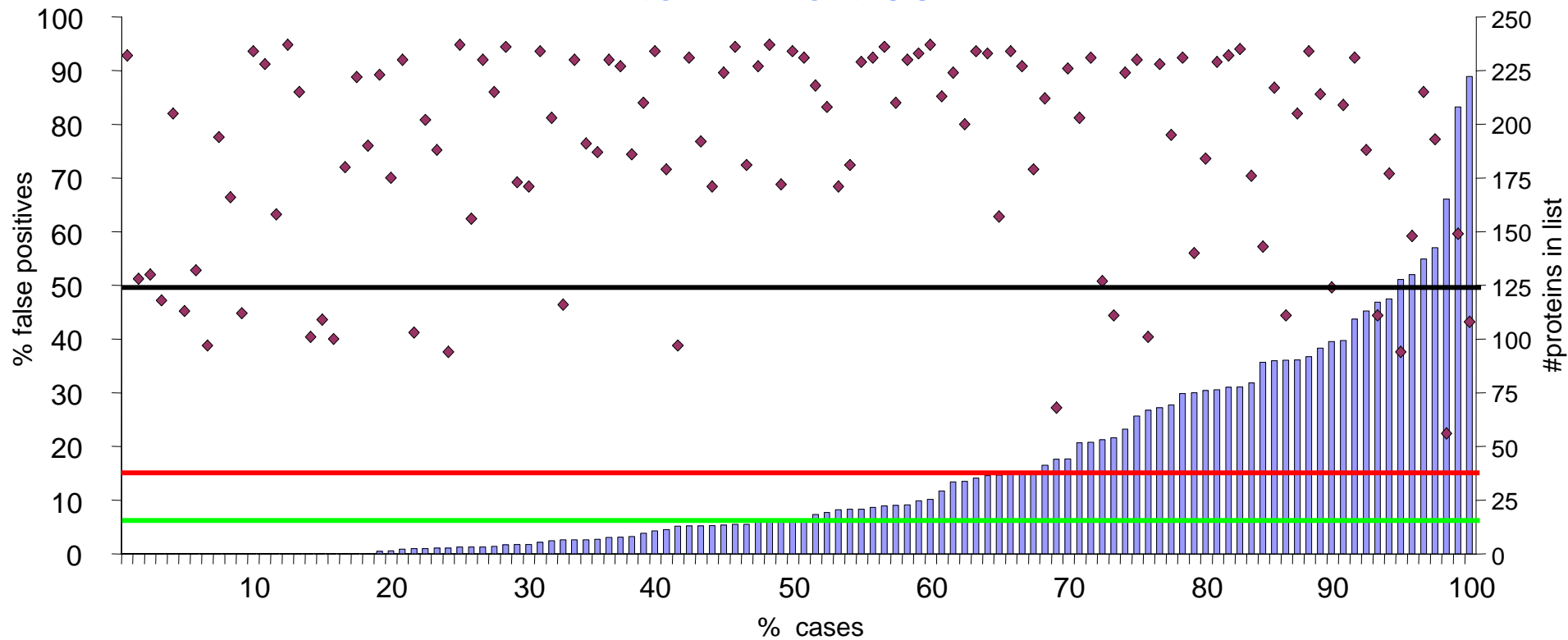


fraction of false positives
0%: perfect
50%: random
(1 int.)



ROC area
1.0: perfect
0.5: random

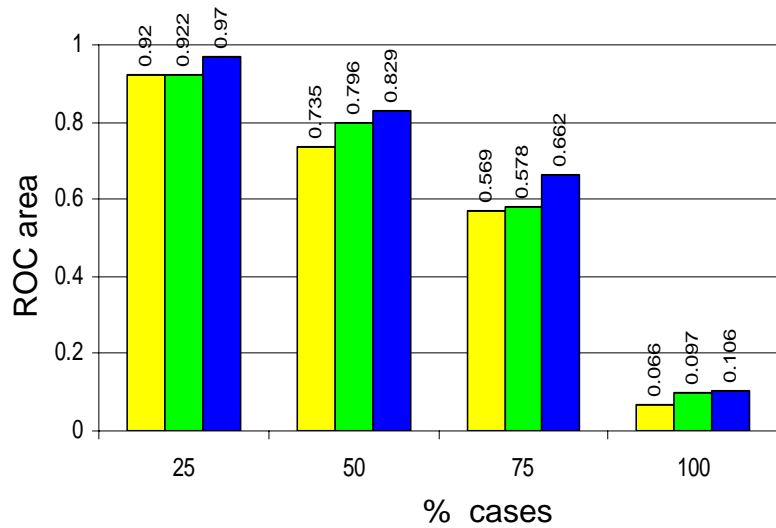
tol-mirrortree



mirrotree

mirrortree
 (tree dist.)

tol-mirrortree



tol-mirrortree

Comparison with *old* methods

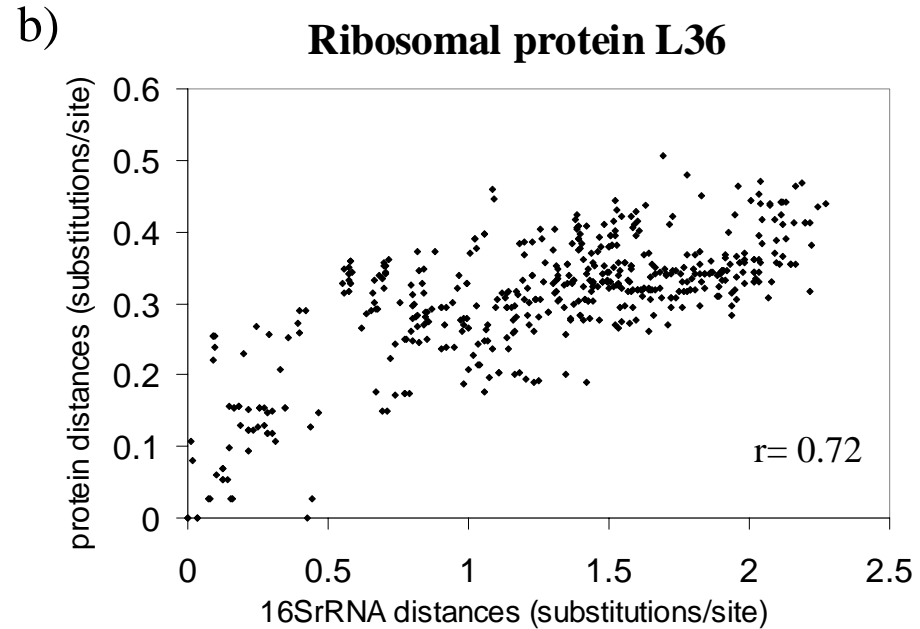
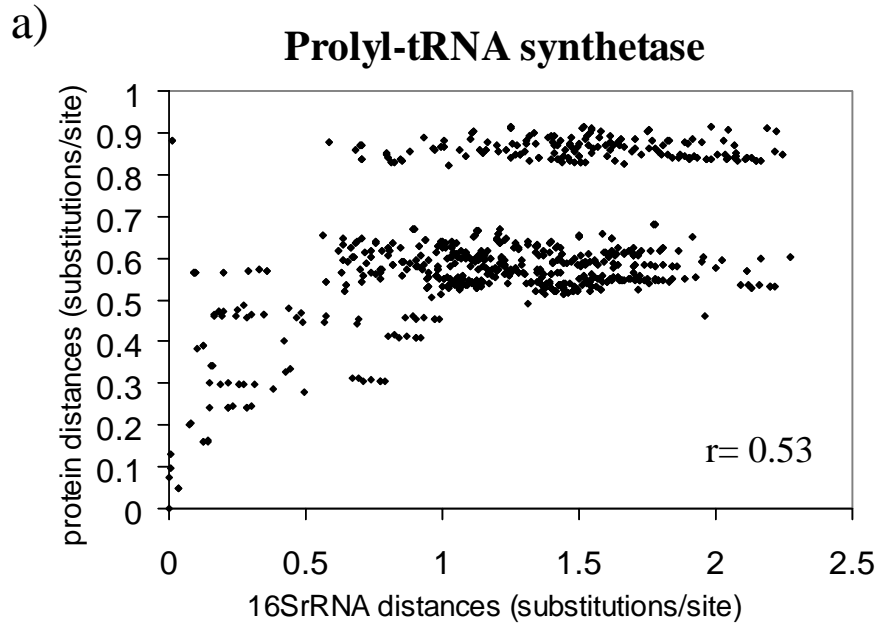
	% false positives	ROC area
<i>mirrortree</i>	23.4	0.71
<i>mirrortree</i> with tree distances	21.9	0.73
<i>tol-mirrortree</i>	14.9	0.79

P(N) values (sign test):

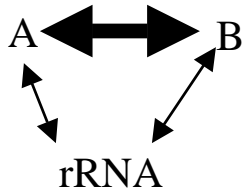
a \ b	<i>mirrortree</i>	<i>mirrortree</i> tree dist.	<i>tol-mirrortree</i>
<i>mirrortree</i>			
<i>mirrortree</i> tree dist.	0.276		
<i>tol-mirrortree</i>	$5.60 \cdot 10^{-6}$	$1.91 \cdot 10^{-5}$	

tol-mirrortree

Prediction of HGT events

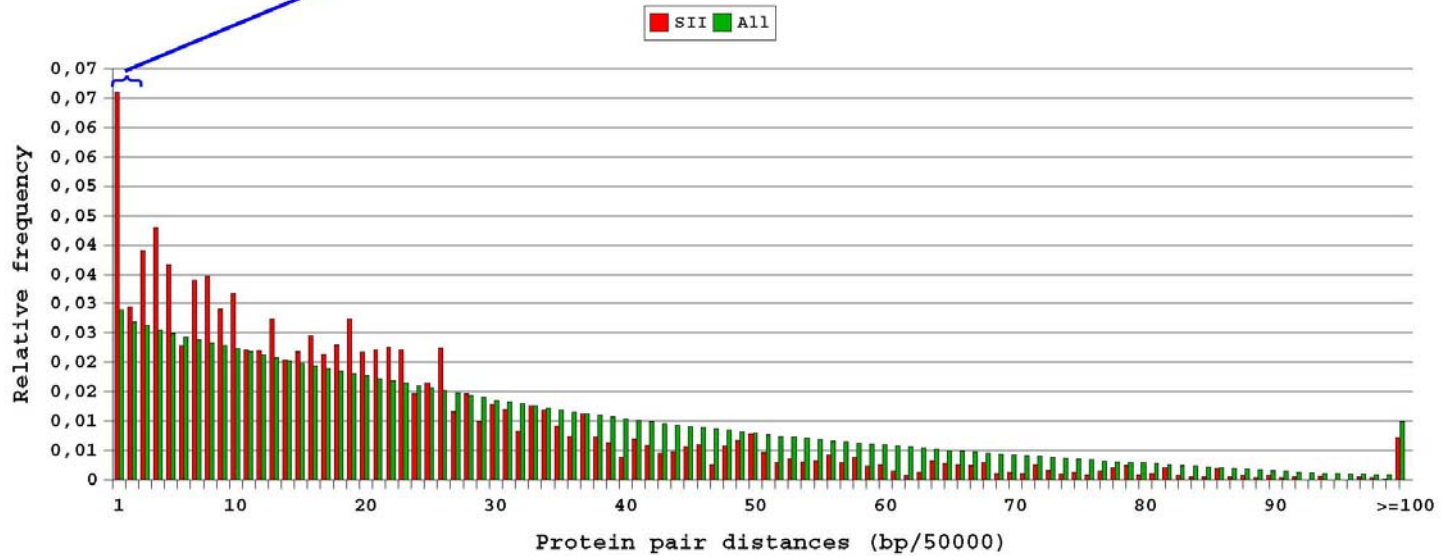
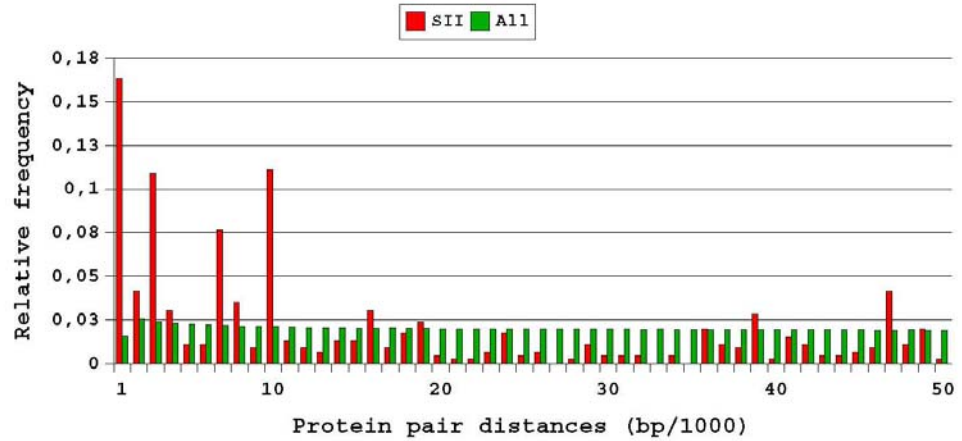


$r \leq 0.5 \rightarrow 25\%$ false pos (vs. 15%)
Excluding them: 13.7% false pos (vs. 15%)



tol-mirrortree Co-HGT events

SII vs All *E. coli* protein orthologous pair distances



Public repositories of protein interactions

ORGANISM	PROTEINS	INTERACTIONS	#
<i>Saccharomyces cerevisiae</i> (budding yeast)	4699	14722	
<i>Helicobacter pylori</i>	710	1415	5
<i>Homo sapiens</i> (Human)	687	714	5
<i>Mus musculus</i> (house mouse)	177	97	
Others	534	714	

What it does ...

STRING is a database of known and predicted protein-protein interactions. The interactions include direct (physical) and indirect (functional) associations; they are derived from four sources:

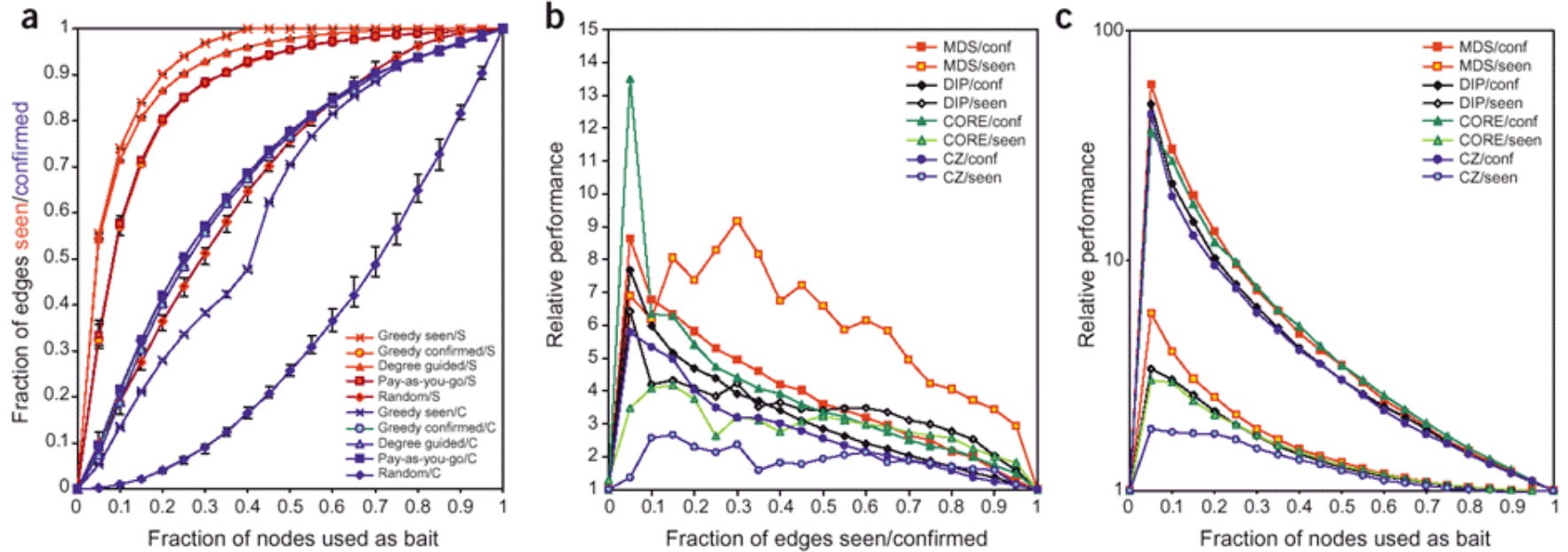
- Genomic Context
- High-throughput Experiments
- (Conserved) Coexpression
- Previous Knowledge

STRING quantitatively integrates interaction data from these sources for a large number of organisms, and transfers information between these organisms where applicable. The database currently contains 736429 proteins in 179 species.

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. & Eisenberg, D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res*, **30**, 303-305.

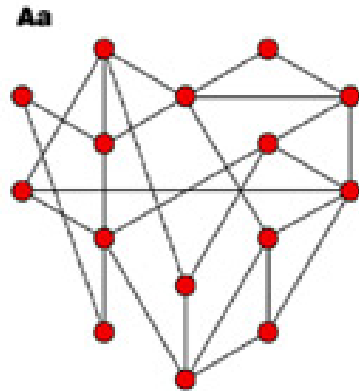
von Mering, C., Huynen, M., Jaeggi, D., Schmidt, S., Bork, P. and Snel, B. (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res*, **31**, 258-261.

Bioinformatics for designing high-throughput experiments

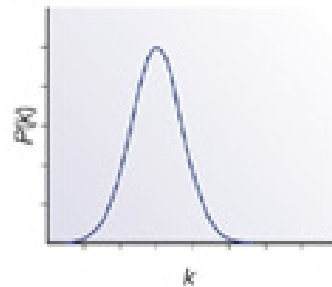


Topological Characteristics of Biological Networks

A Random network



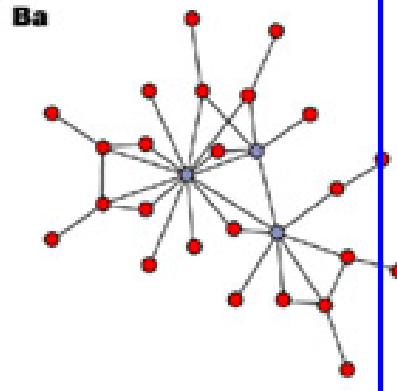
Ab



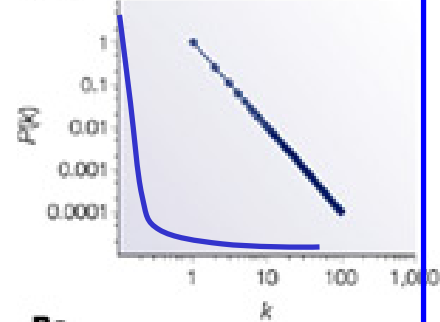
Ac



B Scale-free network



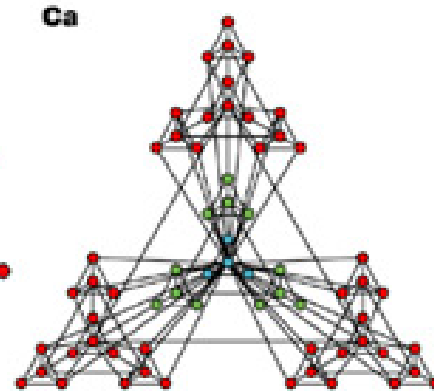
Bb



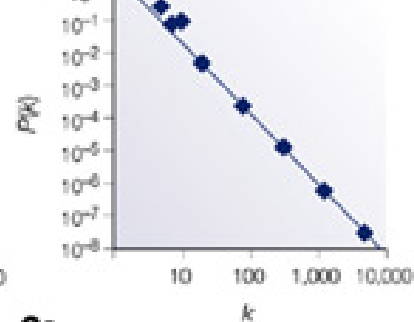
Bc



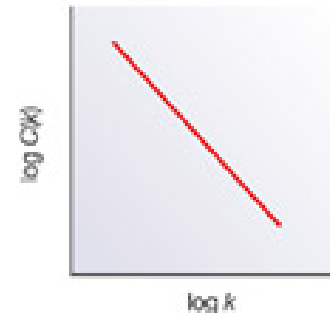
C Hierarchical network



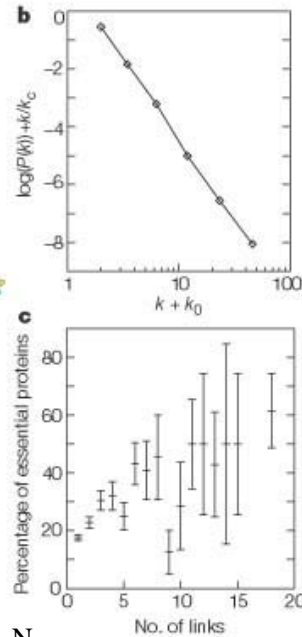
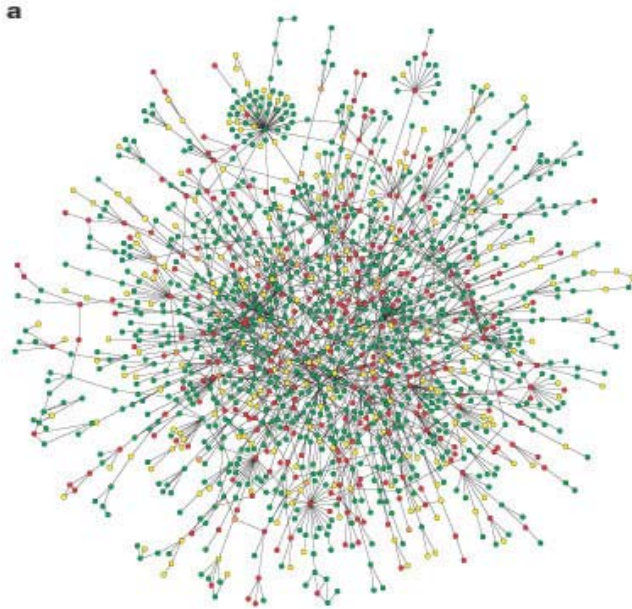
Cb



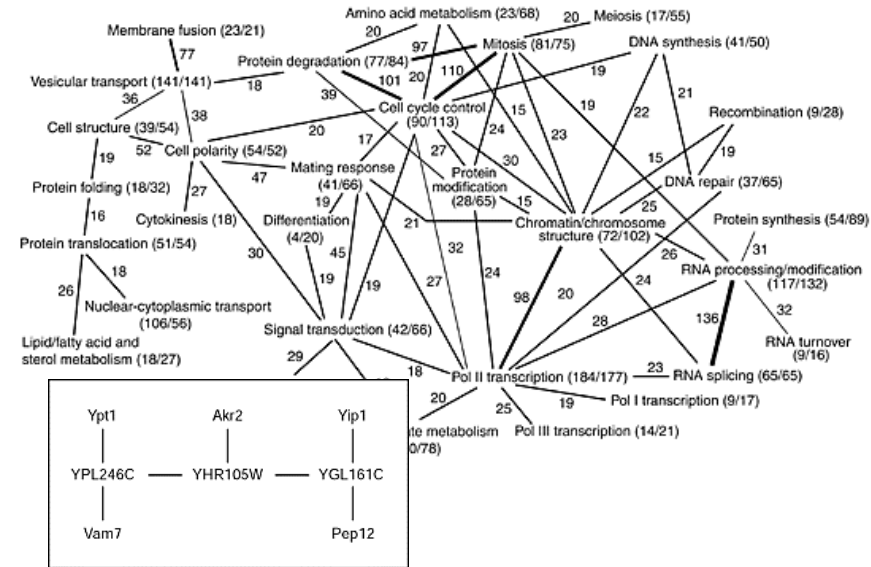
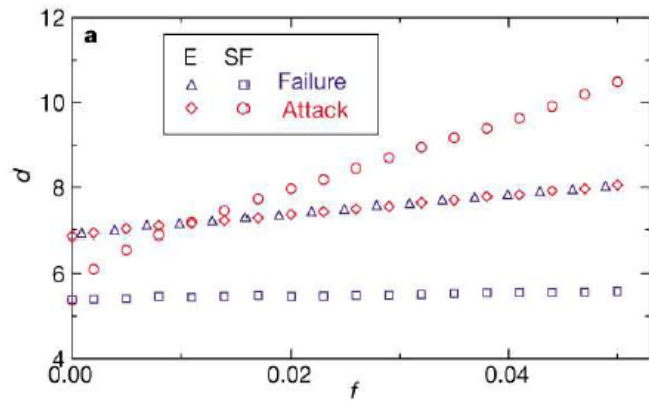
Cc



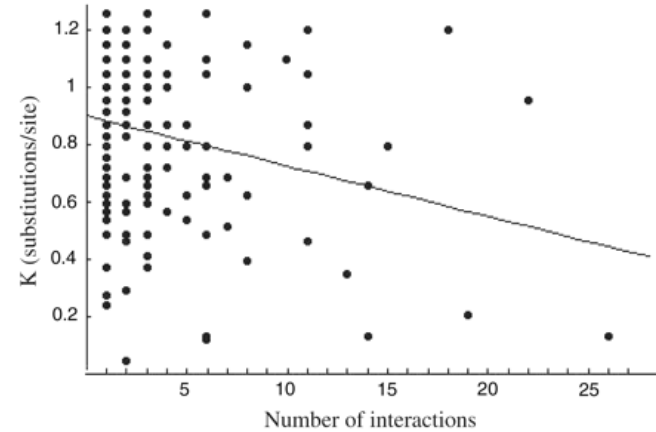
Study of the Interactome



Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. (2001). Lethality and centrality in protein networks. *Nature* **411**, 41-42.

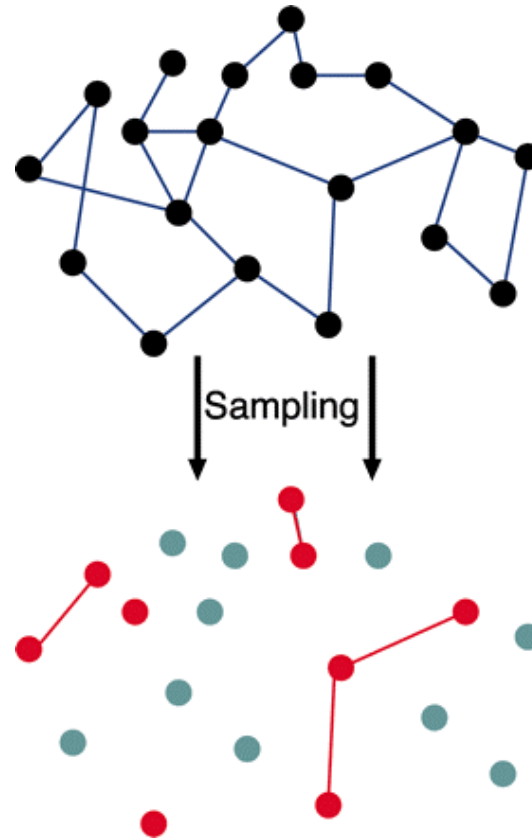


Schwikowski, B., Uetz, P. & Fields, S. (2002). A network of protein-protein interactions in yeast. *Nature Biotech* **18**, 1257-1261.



Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. and Feldman, M.W. (2002) Evolutionary rate in the protein interaction network. *Science*, **296**, 750-752.

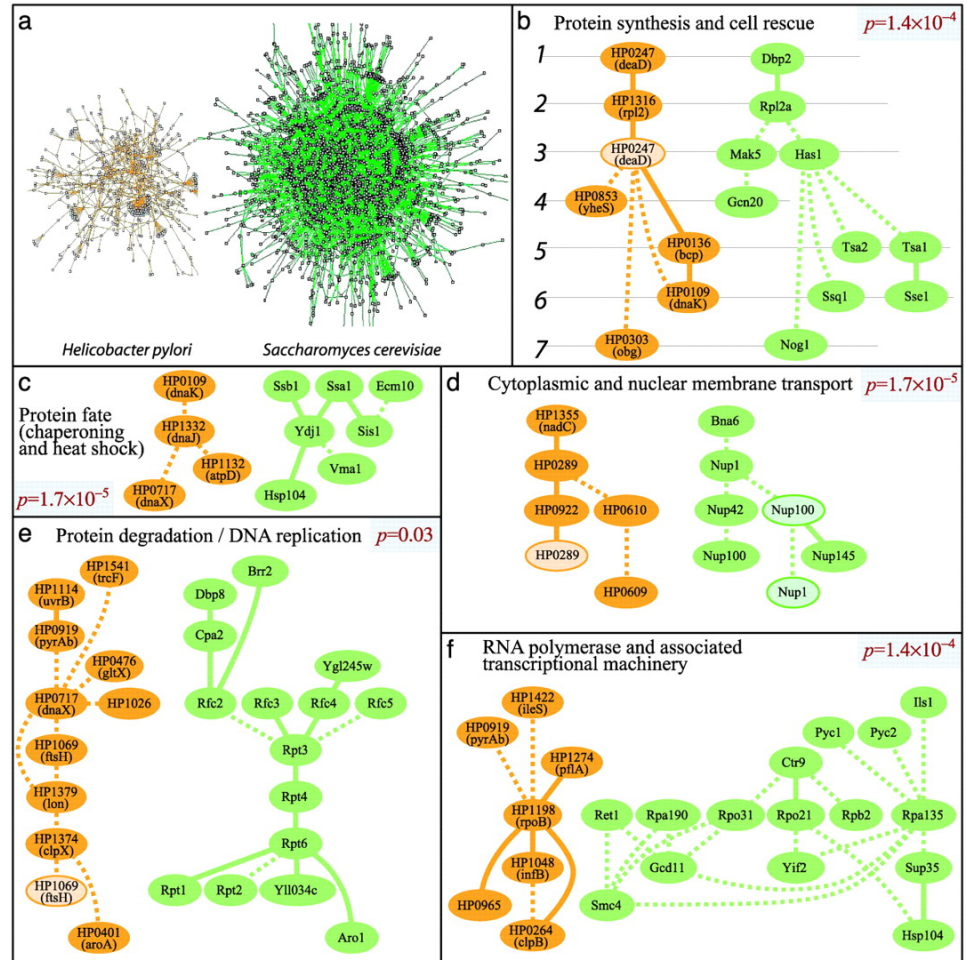
Artifacts due to the sampling process ?



Study of the Interactome

Table 1 Evolutionary conservation of motif constituents

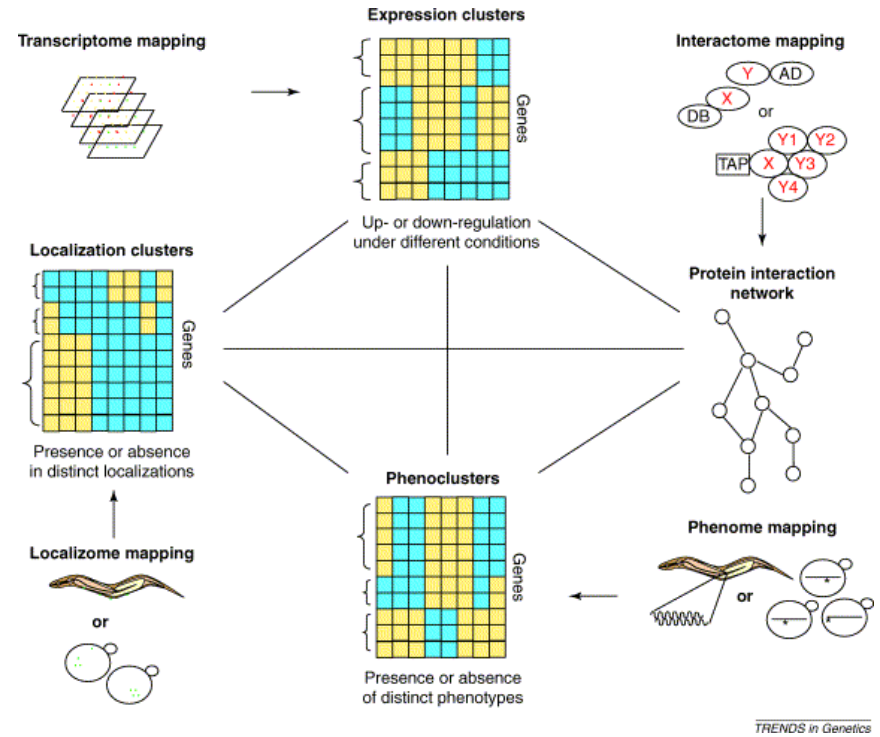
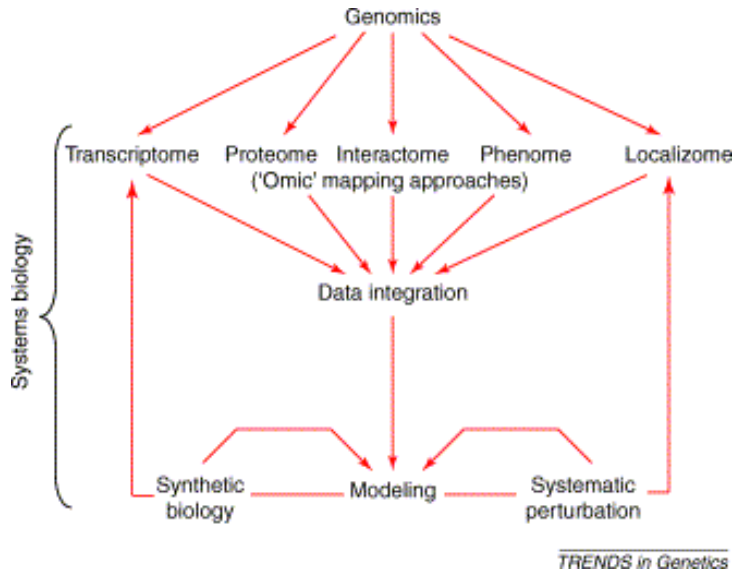
#	Motifs	Number of yeast motifs	Natural conservation rate	Random conservation rate	Conservation ratio
1		9,266	13.67%	4.63%	2.94
2		167,304	4.99%	0.81%	6.15
3		3,846	20.51%	1.01%	20.28
4		3,649,591	0.73%	0.12%	5.87
5		1,763,891	2.64%	0.18%	14.67
6		9,646	6.71%	0.17%	40.44
7		164,075	7.67%	0.17%	45.56
8		12,423	18.68%	0.12%	157.89
9		2,339	32.53%	0.08%	422.78
10		25,749	14.77%	0.05%	279.71
11		1,433	47.24%	0.02%	2,256.67



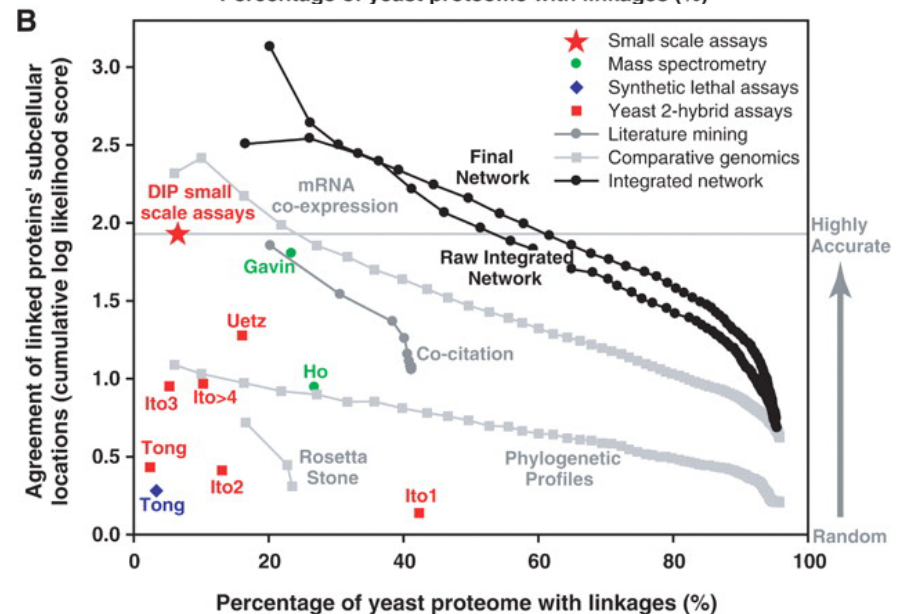
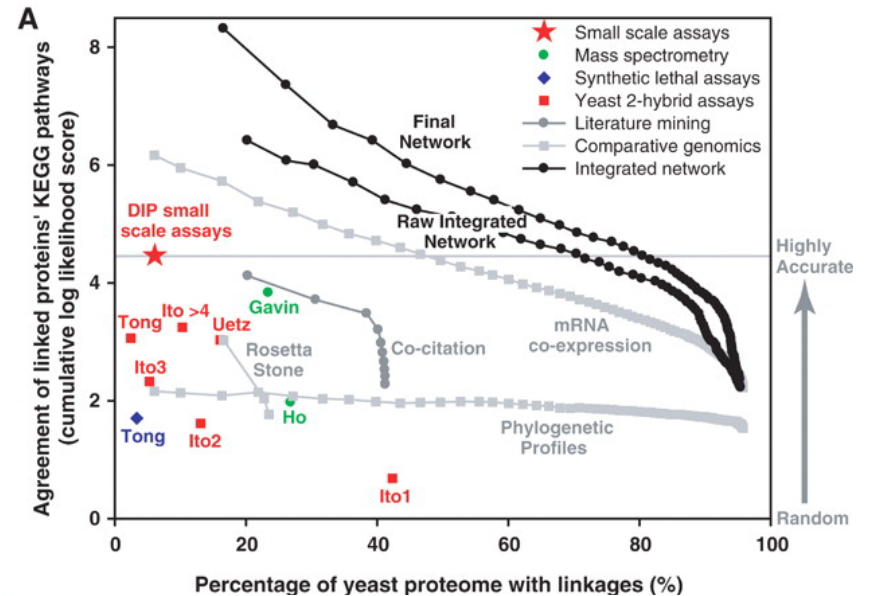
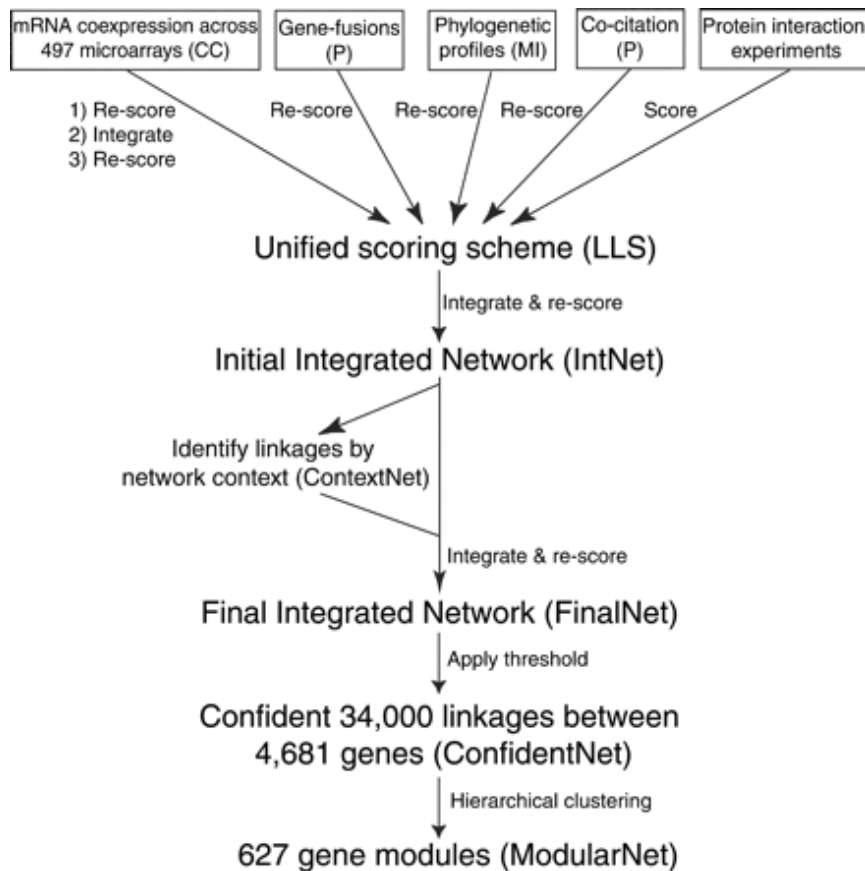
Wuchty, S., Oltvai, Z.N. & Barabasi, A.L. (2003) Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat Genet*, **35**, 176-179.

Kelley, B.P., et al. (2003) Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc Natl Acad Sci U S A*, **100**, 11394-11399.

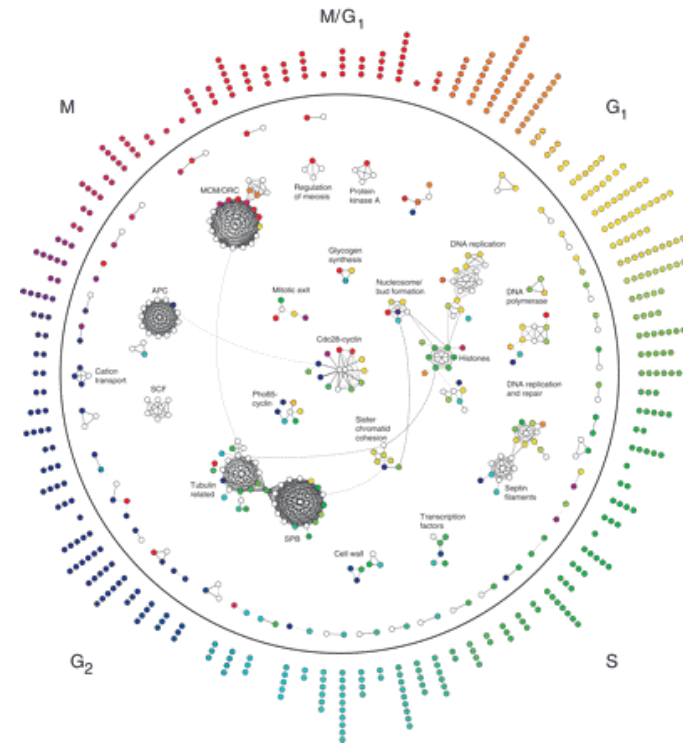
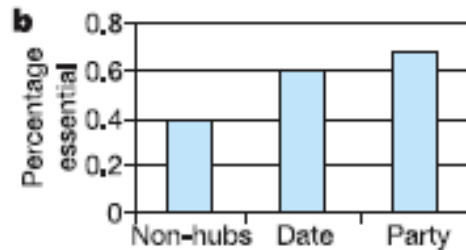
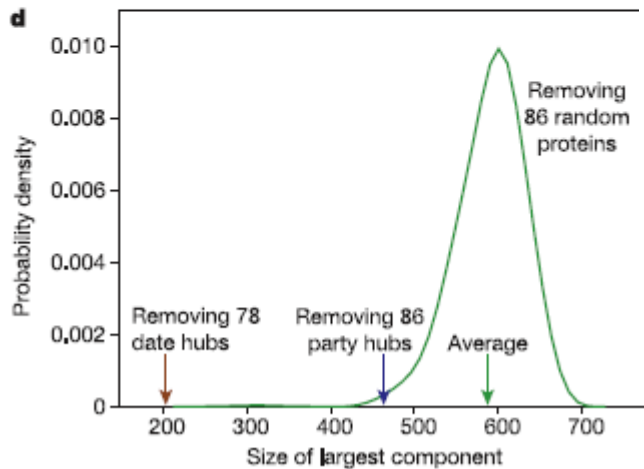
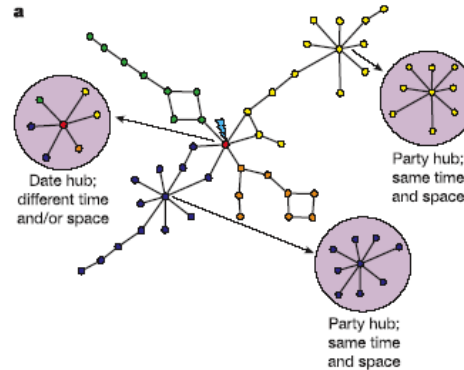
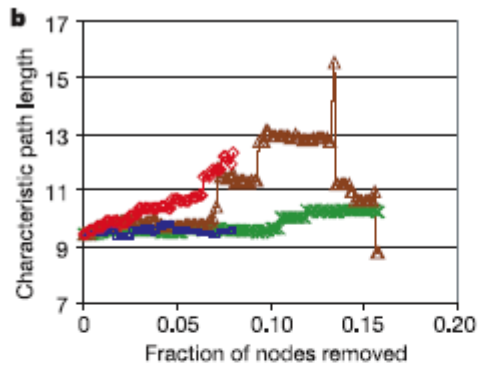
Integration with other “-omics” sources



Combination with other sources to increase reliability



Combination of protein interactions with expression arrays

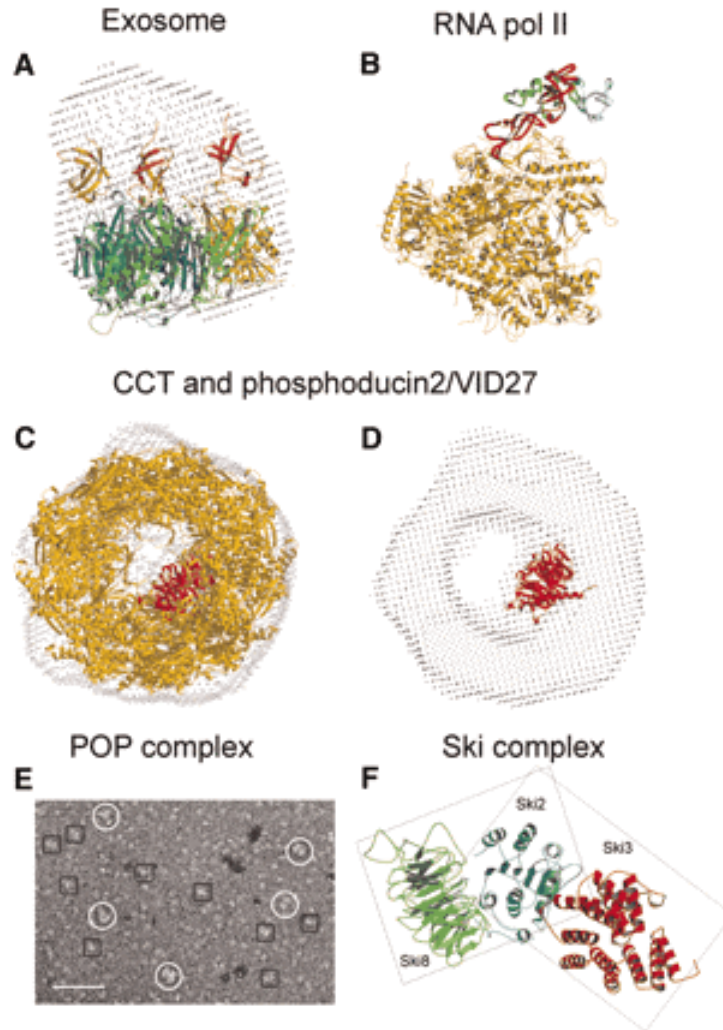


Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., Zhang, L.V., Dupuy, D., Walhout, A.J., Cusick, M.E., Roth, F.P. and Vidal, M. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, **430**, 88-93. Epub 2004 Jun 2009.

de Lichtenberg U, Jensen LJ, Brunak S, Bork P. (2005). Dynamic complex formation during the yeast cell cycle. *Science*. **307(5710)**:724-727.

Going down to the structures

Linking top-down and bottom-up approaches



Redes de Interacciones entre Proteínas

Bibliografía general (revisiones)

- Van Regenmortel, M.H. (2004) Reductionism and complexity in molecular biology. Scientists now have the tools to unravel biological and overcome the limitations of reductionism. *EMBO Rep*, **5**, 1016-1020.
- Uetz, P. and Finley, R.L., Jr. (2005) From protein networks to biological systems. *FEBS Lett*, **579**, 1821-1827.
- Xia, Y., Yu, H., Jansen, R., Seringhaus, M., Baxter, S., Greenbaum, D., Zhao, H. and Gerstein, M. (2004) Analyzing cellular biochemistry in terms of molecular networks. *Annu Rev Biochem*, **73**, 1051-1087.
- Barabasi, A.L. and Oltvai, Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet*, **5**, 101-113.
- Alm, E. and Arkin, A.P. (2003) Biological networks. *Curr Opin Struct Biol*, **13**, 193-202.
- Salwinski, L. and Eisenberg, D. (2003) Computational methods of analysis of protein-protein interactions. *Curr Opin Struct Biol*, **13**, 377-382.

Redes de Interacciones entre Proteínas

Material de teoría y práctica:

http://pdg.cnb.uam.es/pazos/cursos/Verano_UCM_05/

Contacto:

Dr. Florencio Pazos Cabaleiro
Protein Design Group
Centro Nacional de Biotecnología (CNB-CSIC). Madrid
pazos@cnb.uam.es
<http://pdg.cnb.uam.es/pazos>