Biological network models

Approaches used to understand the mechanisms behind the evolution of molecular networks

- Growth/ preferential attachment models: duplication of genes, metabolic reactions, recombination, development
- Comparative genomics/ pathway comparison across phylogenetic tree
- Predicting interactions based on existing information
- Study the evolution of motifs and cliques

Evolving protein interaction networks

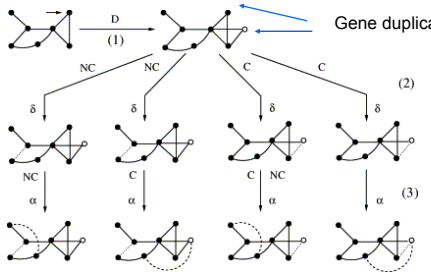
Genes and interactions among gene products have often been conserved through evolution (orthologs).

We can consider the topology of protein interaction networks as a result of a network evolution process.

One can formulate evolving network models for protein interaction networks.

Driving forces behind the formation of edges:
gene duplication and mutation
protein attractiveness (affinity)
– determined by protein (domain) structure

Duplication-divergence models



Gene duplication

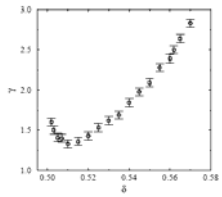
Edge loss with prob. δ

Edge gain with prob. α

Correlated connections (C): only the duplicated gene loses/ gains edge
Uncorrelated connections (NC): edge can be added or removed between any pair of nodes in the network.

Pastor-Satorras et al., Journ. Theor. Biol 22, 199 (2003)

Network properties



Assume size- dependent edge gain constraint

$$\alpha = \beta / N$$

$$P(k) \propto (k_0 + k)^{-\gamma} e^{-\frac{k}{k_c}} \qquad \langle k \rangle = \frac{2\beta}{2\delta - 1}$$

$$k_c = (\ln(\delta/\delta - 1))^{-1}$$

Apart from a concave region γ is increasing function of δ.

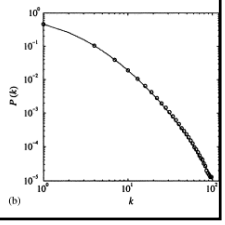
$$\gamma = -k_0 = 1 - \frac{2\beta}{1-\delta}$$

Good agreement with yeast prot. int. network (of N=2000, <K> = 2.5) if δ=0.562

$\gamma = 2.5 \pm 0.1$

$K_c \sim 37$

The clustering coefficient of a gene duplication model depends strongly on the initial seed network on which the duplication is performed



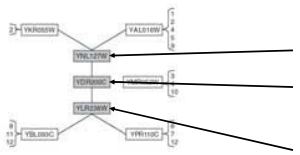
Comparison of networks between different species

- Assembly of cellular networks:
  - Integrate different kinds of interactions
  - Predict new interactions by integrating indirect information e.g. co-expression, co-citation, sequence similarity
- Use interaction information to predict gene function or functional similarity (orthology) between genes
- Use networks from different organisms to find functionally conserved genes, interactions and network motifs

Functionality of proteins

- Approaches to predict functions – clustering of co-regulated genes, phylogenetic profiles, protein complexes.
- Assignment of functional classes on the basis of their network of physical interactions.

Subgraph of protein interaction network.



Functional classification

(2) – budding, cell polarity, filament formation

(3, 4, 10) – pheromone response, sex-specific proteins, cell cycle

(12) – nuclear organization

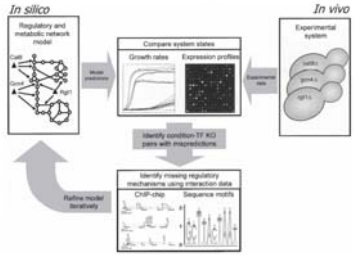
Lee, et. al. 2004 Science (306) 1555-58
Vazquez, et al. 2003 Nat. Biotech. (21) 697-700

Integrated analysis of regulatory and metabolic networks

Regulatory network - Boolean
Metabolic network – flux balance analysis
1. activity of TF in response to extra- and intracellular metabolite concentrations
2. expression of metabolic genes in response to TFs and metabolites.

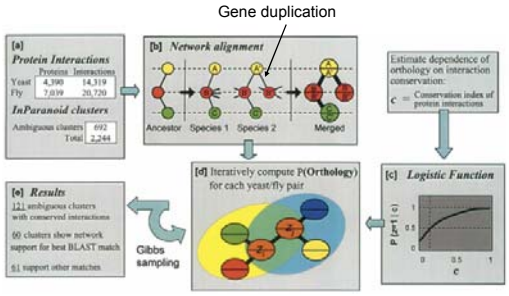
Network expansion – regulatory cascade connecting TF to genes encoding enzymes using transcription factor binding data.

Predicts growth phenotypes of TF knockouts
The expanded network predicts previously uncharacterized interactions between TFs.



Herrgard, et. al. 2006 Genome Res. 16: 627-635

Network alignment of yeast and fly protein-protein interaction networks to find functional orthologs



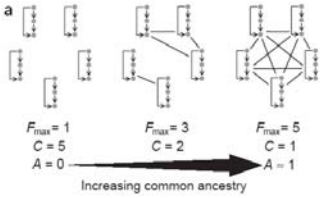
Only conserved interactions are used.
The network supports functional orthology for 61 pairs that are not the most conserved.

Bandyopadhyay, et. al. 2006 Genome Research 16: 428-35

Evolutionary origins of cellular network motifs

- The motifs may have evolved through random duplication and subsequent diversification of ancestral circuits
- Convergent evolution - These motifs can arise from unrelated genes

Indicators of evolutionary origin of circuits:
1. Index of common circuit ancestry A
2. Size of largest family of circuits with common ancestry F_{max}



Conant, et. al. 2003 Nat. Genet. 34: 264-266

Optimal circuit design

Motifs in *S. cerevisiae* and *E. coli*.

E. coli – Bi-fans, feed forward loops

S. cerevisiae – Bi- fans, feed forward loops, multi-input motifs, regulatory chains.

- No evidence for common ancestry of *E. coli* motifs.
- No common ancestry of yeast regulatory chains, feed-forward loops and multi-input motifs with more than 2 regulators.
- Yeast bi-fan motifs show some common ancestry.

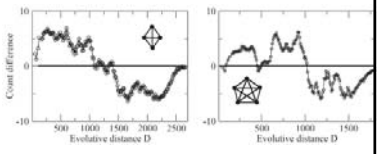
b

	Circuit type	Number of circuits	Number of families (C)	Index of common ancestry (A)	Largest circuit family (F_{max})
Yeast	Feed-forward	48	44 (46.8 ± 1.9; P = 0.08)	0.082 (0.023 ± 0.035; P = 0.08)	5 (1.9 ± 1.4; P = 0.06)
	Bi-fan	542	435 (469.0 ± 37.7; P = 0.18)	0.197 (0.135 ± 0.070; P = 0.18)	49 (41.0 ± 31.1; P = 0.33)
	MIM-2	176	168 (164.5 ± 8.8; P = 0.60)	0.045 (0.065 ± 0.050; P =0.60)	5 (7.4 ± 6.2; P = 0.59)
	Reg. chain (3)	33	33	0	1
E. coli	Feed-forward	11	11	0	1
	Bi-fan	27	27	0	1

Co-operative co-evolution

- Comparative analysis of the protein interaction networks of four species of the class hemiascomycetes (that includes *S. cerevisiae)* .
- Hypothesis: proteins in cliques evolve together, suggesting that compensatory mutations take place.
- Evolutionary divergence rates are calculated from the sequence similarity of *S. cerevisiae* proteins participating in the cliques with the other species.

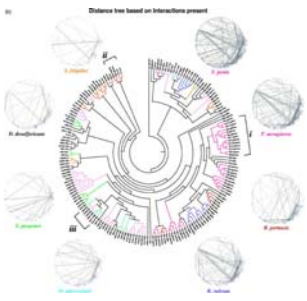
Shown: count difference in real versus randomized networks.



Co-evolving cliques are abundant.

Vergassola, et. al. 2005 Proteomics 5: 3116-19

Conservation of regulatory interactions in different genomes



Reconstructed gene regulatory networks based on E. coli network and orthology relationships

Hierarchically clustered genes based on their interaction conservation profiles

i – same phylogenetic group
ii – parasites
iii – similar lifestyle but different phylogenetic group

Different TF hubs evolve independently – convergent evolution

Madan Babu, et. al. 2006 J. Mol. Bio. 358: 614-633