

Conserved ancestral protein-protein interactions

Janusz Dutkowski
and
Jerzy Tiuryn

Institute of Informatics, Warsaw University



May 28, 2007

Outline

- Protein interaction networks.
- Network alignment problem.
- Earlier approaches:
 - Kelley, Sharan, Ideker 2003–2005,
 - Koyutürk 2005.
- Ancestral network reconstruction.
- Model of protein network evolution.
- Ancestral module detection.
- Experiments and results.

Proteins

- **Proteins** are made of amino acids arranged in a linear chains.
- Proteins fold into unique 3D structures (native state).



Figure: Phillips, *J. Mol. Biol.*,1980.

- Most cellural functions are realized by interacting proteins.

Protein interaction networks

- Graph $G = (V, E)$.
- Nodes represent proteins.
- Edges represent physical protein interactions.
- Optional: edge weights denote probability of interaction.

Network alignment problem

Detecting similar/dissimilar regions in a set of networks:

- similar **protein sequences**
- similar **interaction patterns**
- Motivation: similarity (**conservation**) implies **functionality**
 - conserved clusters indicate complexes,
 - conserved linear paths indicate signaling pathways.

Network alignment problem

Detecting similar/dissimilar regions in a set of networks:

- similar **protein sequences**
- similar **interaction patterns**
- Motivation: similarity (**conservation**) implies **functionality**
 - conserved clusters indicate complexes,
 - conserved linear paths indicate signaling pathways.
- **Computationally hard** (largest common subgraph, subgraph isomorphism).

- 1 Construct network alignment graph:
 - Nodes represent sets of potentially homologous proteins (BLAST E value), one from each network.
 - Edges represent conserved interactions across the aligned networks.

Previous approaches – general scheme

- 1 Construct network alignment graph:
 - Nodes represent sets of potentially homologous proteins (BLAST E value), one from each network.
 - Edges represent conserved interactions across the aligned networks.
- 2 Scoring function measures conservation of protein sequences and interactions.

Previous approaches – general scheme

- 1 Construct network alignment graph:
 - Nodes represent sets of potentially homologous proteins (BLAST E value), one from each network.
 - Edges represent conserved interactions across the aligned networks.
- 2 Scoring function measures conservation of protein sequences and interactions.
- 3 Greedy algorithm for identifying conserved subnetworks of a certain structure.

Previous approaches – general scheme

- 1 Construct network alignment graph:
 - Nodes represent sets of potentially homologous proteins (BLAST E value), one from each network.
 - Edges represent conserved interactions across the aligned networks.
- 2 Scoring function measures conservation of protein sequences and interactions.
- 3 Greedy algorithm for identifying conserved subnetworks of a certain structure.
- 4 Search for high-scoring linear paths in alignment graph (PathBLAST).
- 5 Search for conserved dense subgraphs in two interaction networks (NetworkBLAST, MaWISh).

What was there in the beginning?

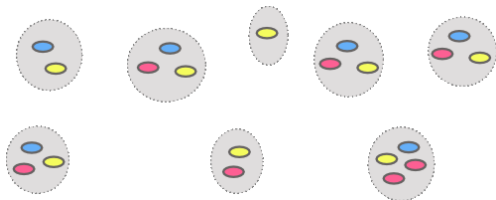
- We are looking for modules of **conserved** interactions.
- **Conserved** interactions imply the existence of ancestral interactions.
- **Approach:** Find **ancestral interactions** with strong support in the input networks.

Conserved Ancestral Protein-Protein Interactions

1) Input PPI networks

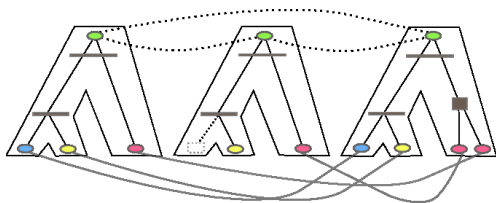


2) MCL clustering of proteins based on sequence distances



3) Build reconciled gene trees

4) Compute the probability of each ancestral interaction given the observed data and model of network evolution.



Dutkowski and Tiuryn, Bioinformatics 2007.

General duplication and divergence model

Parameters: p_d , δ_d , p_s , and δ_s .

- Start with the ancestral graph and perform a defined sequence of duplications and speciations^a.
- In case of **duplication** of protein p replace it by its two copies a and b . And perform divergence steps:
 - ① for each copy retain each of its interactions with probability p_d , and
 - ② add non-existent interactions, each with probability $\delta_d/|V_t|$.
- In case of **speciation** make two copies of the graph and perform the following divergence steps independently:
 - ① retain each interaction with probability p_s , and
 - ② add each non-existent interaction with probability $\delta_s/|E_t|$.

where V_t and E_t are the sets of vertices and edges in the respective graphs.

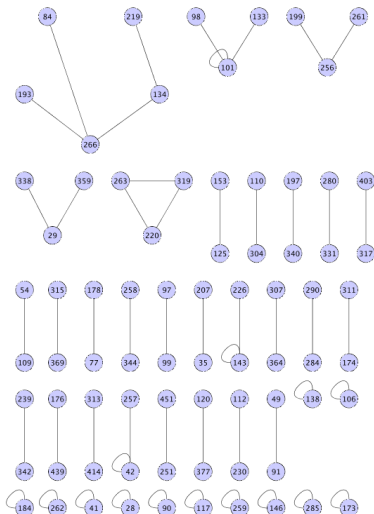
^aThe sequence of events extracted from the reconciled gene trees

Experiments

- PPI data of yeast, worm and fly from DIP database.
- Sequence clustering: 460 of 6971 clusters had at least one representative from each of the three species.
- For each of the 460 clusters (protein families) a gene tree was constructed and reconciled with the species tree.
- Posterior probabilities of pairwise interactions between the 460 ancestral nodes were computed.
- FDR edge weight q -values were calculated using randomized networks.
- Ancestral edges with weights ≤ 0.48 (q -value of 0.049) were discarded. We also removed nodes without any interactions.

Ancestral modules

- Ancestral network decomposed by eliminating edges below the threshold value of 0.48.
- 40 connected components (modules) containing 75 ancestral nodes identified.



Results

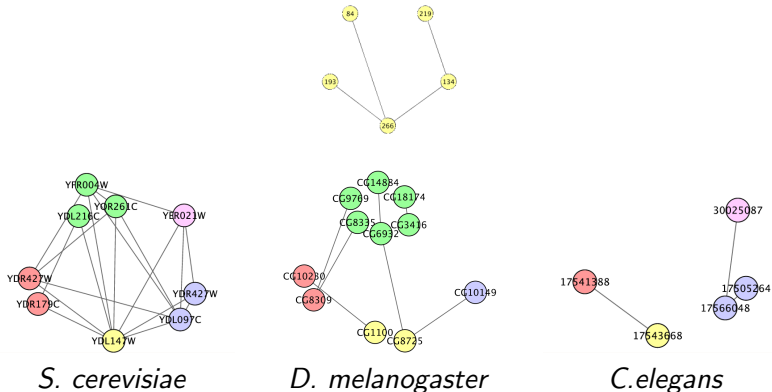


Figure: Projection of ancestral module 193-266-134-219-84 onto the networks of yeast, fly and worm.

Results

Module	Annotated proteins	Purity	MIPS category	Description
193 - 266 - 134 - 219 - 84	7	1	360.10.20	Proteasome
28	14	1	360.10.10	Proteasome
257 - 42	6	0.83	410.40.30	Replication
311 - 174	5	1	500.40.10	Translation
331 - 280	3	1	500.20.10	Translation
176 - 439	4	1	510.190.110	Transcription
176 - 439	4	1	510.190.40	Transcription
41	3	0.67	510.190.130	Transcription
199 - 256 - 261	4	0.5	510.70.20	Transcription
199 - 256 - 261	4	0.5	510.190.10	Transcription
153 - 125	4	1	260.50.20	Intracellular transport
199 - 256 - 261	4	0.5	230.20.20	Histone acetyltransferase

Table: Pure modules and respective annotations from the MIPS database

References

- Kelley et al. Conserved pathways within bacteria and yeast as revealed by global protein network alignment. PNAS, 2003.
- Sharan et al. Conserved patterns of protein interaction in multiple species. PNAS, 2005.
- Koyuturk et al. Pairwise alignment of protein interaction networks. JCB, 2006.
- Flannick et al. Graemlin: general and robust alignment of multiple large interaction networks. Genome Res, 2006.
- Sharan et al. Modeling cellular machinery through biological network comparison. Nature Biotechnology, 2006.
- Dutkowski and Tiuryn. Identification of functional modules from conserved ancestral protein–protein interactions. Bioinformatics, 2007 (proceedings of ISMB/ECCB).

Thank you!