The topology and evolution of a network model for genome evolution in polyploids

PHYS 597A

IBIOS Graduate Program, BG Option
The Huck Institutes for Life Sciences

Gue Su Chang

Outline

- Background:
  Polyploidy and gene divergence

- Gene-Protein Contact (GPC) model
- GPC model: Rate equation
- GPC model: Result

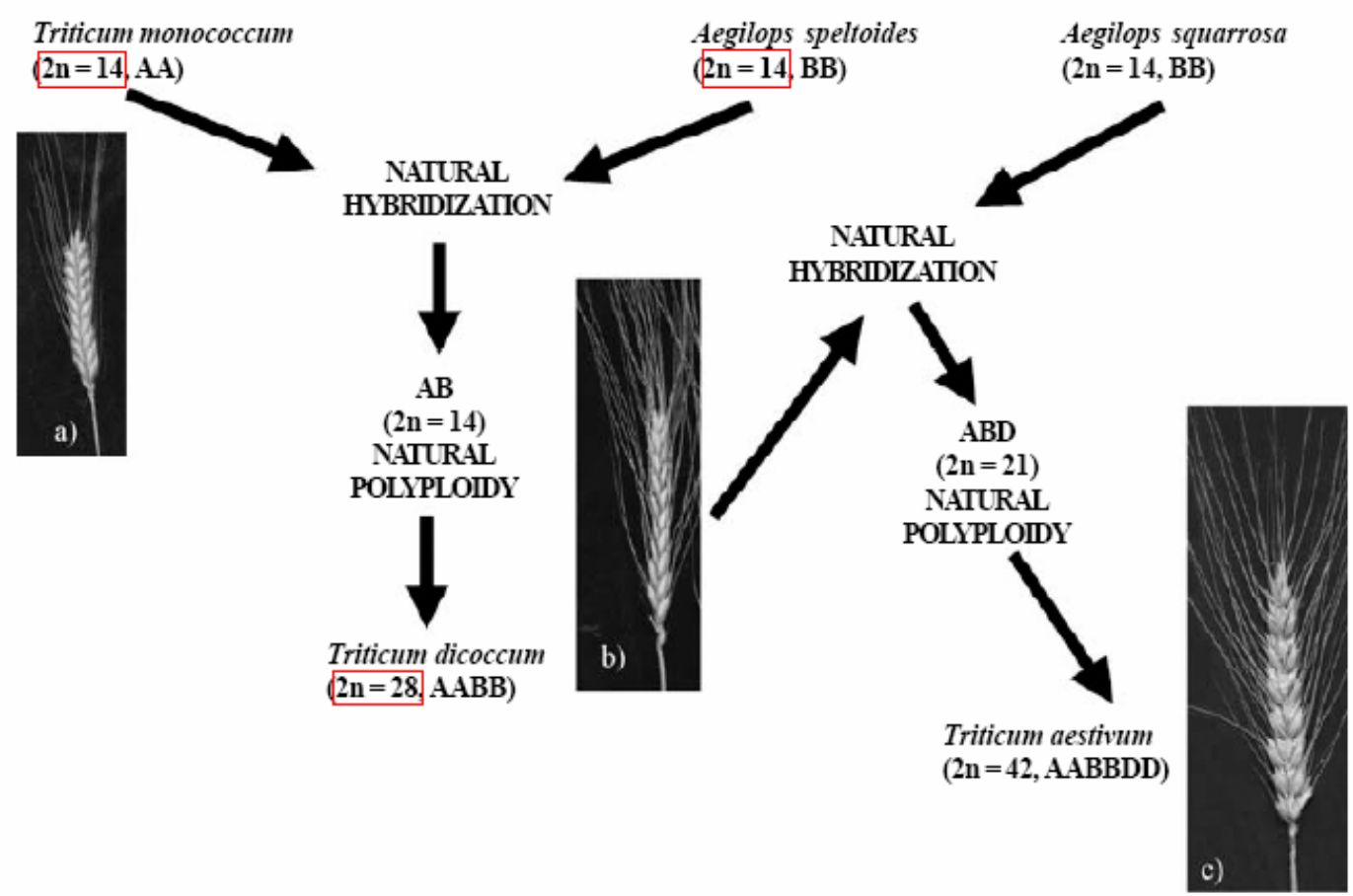
- Discussion
- References

Polyploidy and divergence

- Polyploid
  - Of a nucleus, having more than two haploid sets (*2n*) of chromosomes, for example, 3n is *triploid*, and 4n, *tetraploid*.
  - Autopolyploids (A+A=2AA) and allopolyploids (A+B=2AB)

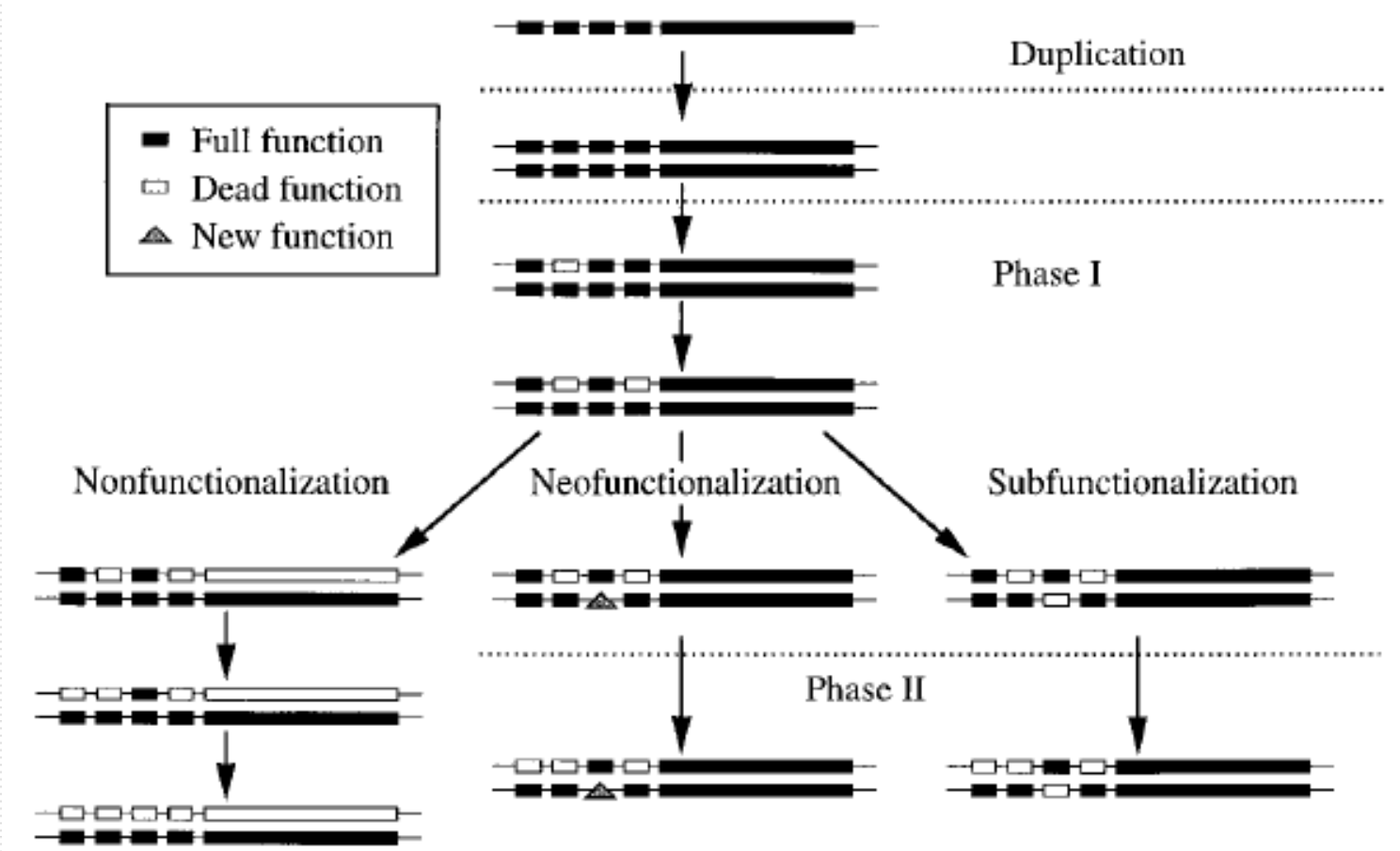
- Polyploidization
  - Gene duplication found in the evolutionary history of 50 ~ 70% of angiosperms
  - Important role in the evolution of diversity and novel functions of genes

Prevalent in genomes of flowering plants



Synoptic chart of cultivated wheat evolution
(de Moraes Fernandes, et al. 2000)

The fate of duplicate gene pairs



Three potential fates of duplicate gene pairs with multiple regulatory region (Force et al., 1999)

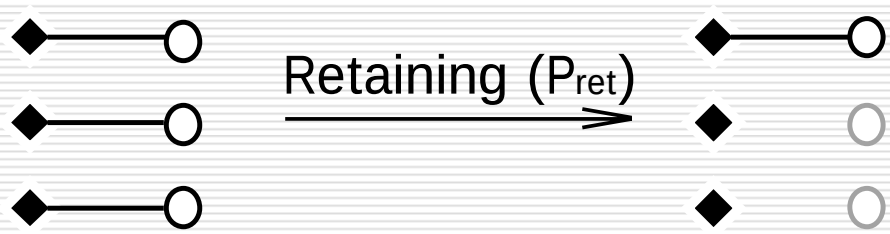
Gene-Protein Contact model

- Assumption

- Considering only divergence in sequence of genes, but not of regulatory regions.
- P_{mut} is the probability of mutation affecting gene function.
- A link is the transcription relation between a gene and its product, protein.
- Not considering the time scale of duplication and divergence of genes. (considering as the steps of the evolution process)

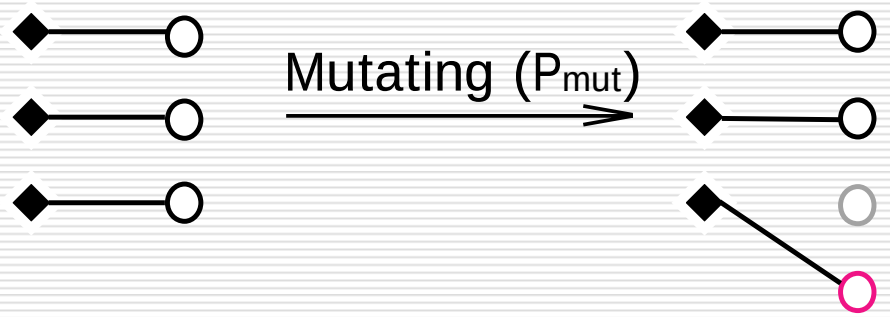
GPC model: Gene divergence

- Retaining and gene silencing



* Fixing as $P_{ret}=0.5$ from Ohno (1970)

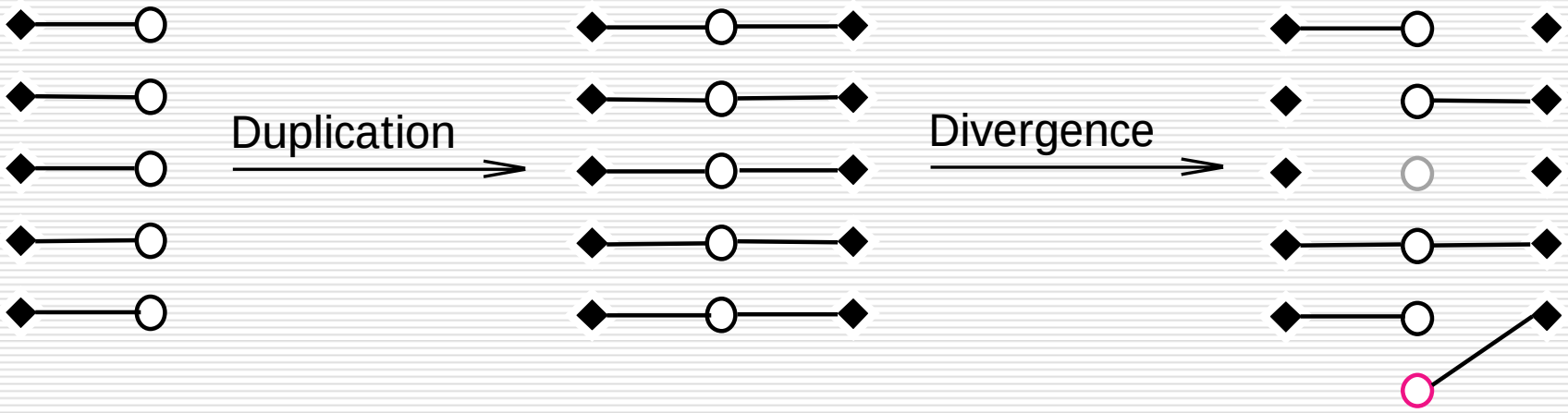
- Acquiring a new function



◆ Gene, O Protein, O New protein

GPC model: Evolving network

- Genome evolution via duplication (n -> 2n*) and divergence



* Starting with the haploid number of chromosomes, $N_0 = 100$

◆ Gene, O Protein, O New protein

GPC model: Rate equation

- The expected number of nodes with *k* links at step *t*

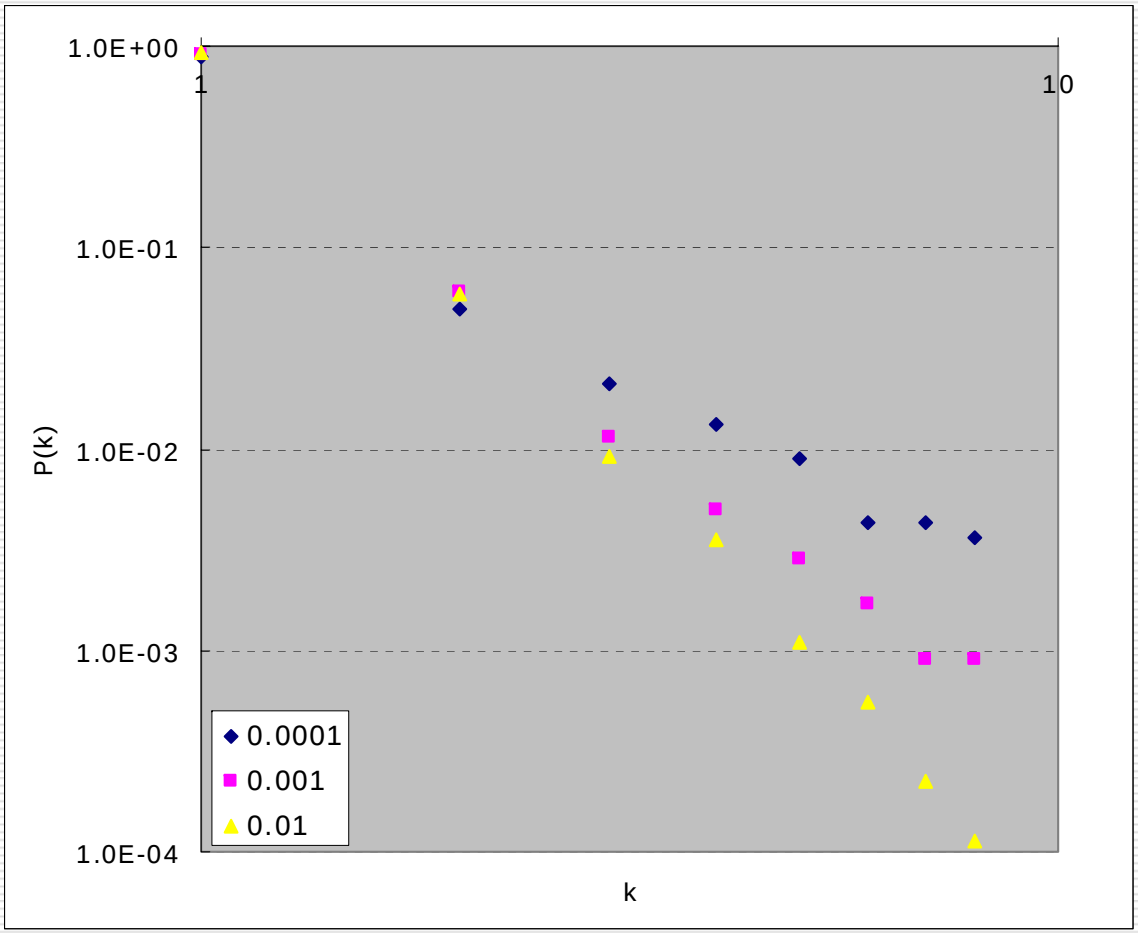
$$f(k,t) = N_t P(k,t)$$

$$f(k,t+1) = 2 \times f(k,t) - P_{mut} \frac{f(k,t)}{N_t}$$

N_t : The total number of nodes at step t

P_{mut} : The probability of mutation

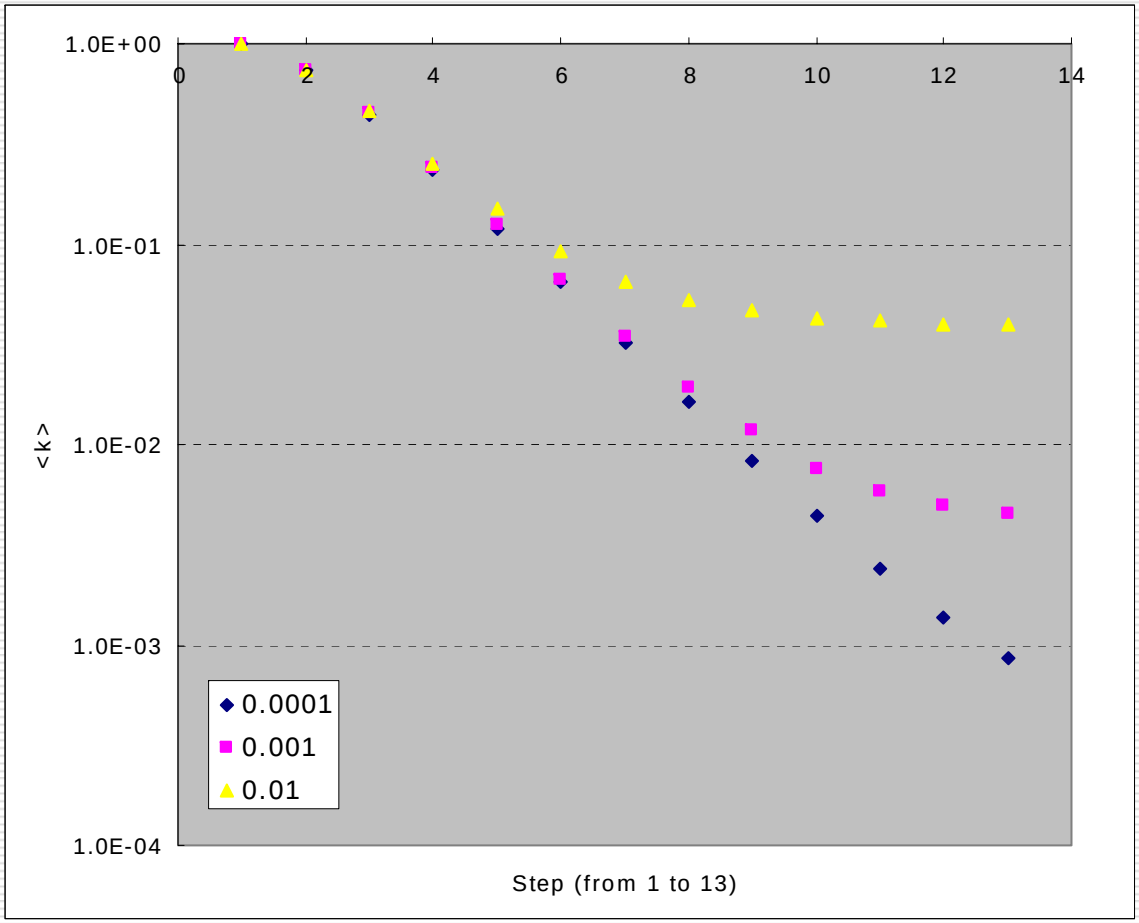
GPC model: Cumulative degree distribution



Degree distributions with different mutation rates (P_{mut} = 0.01, 0.001, 0.0001)

* Data shown are means over 10 graphs with of 400,000 nodes

GPC model: Average degree versus step

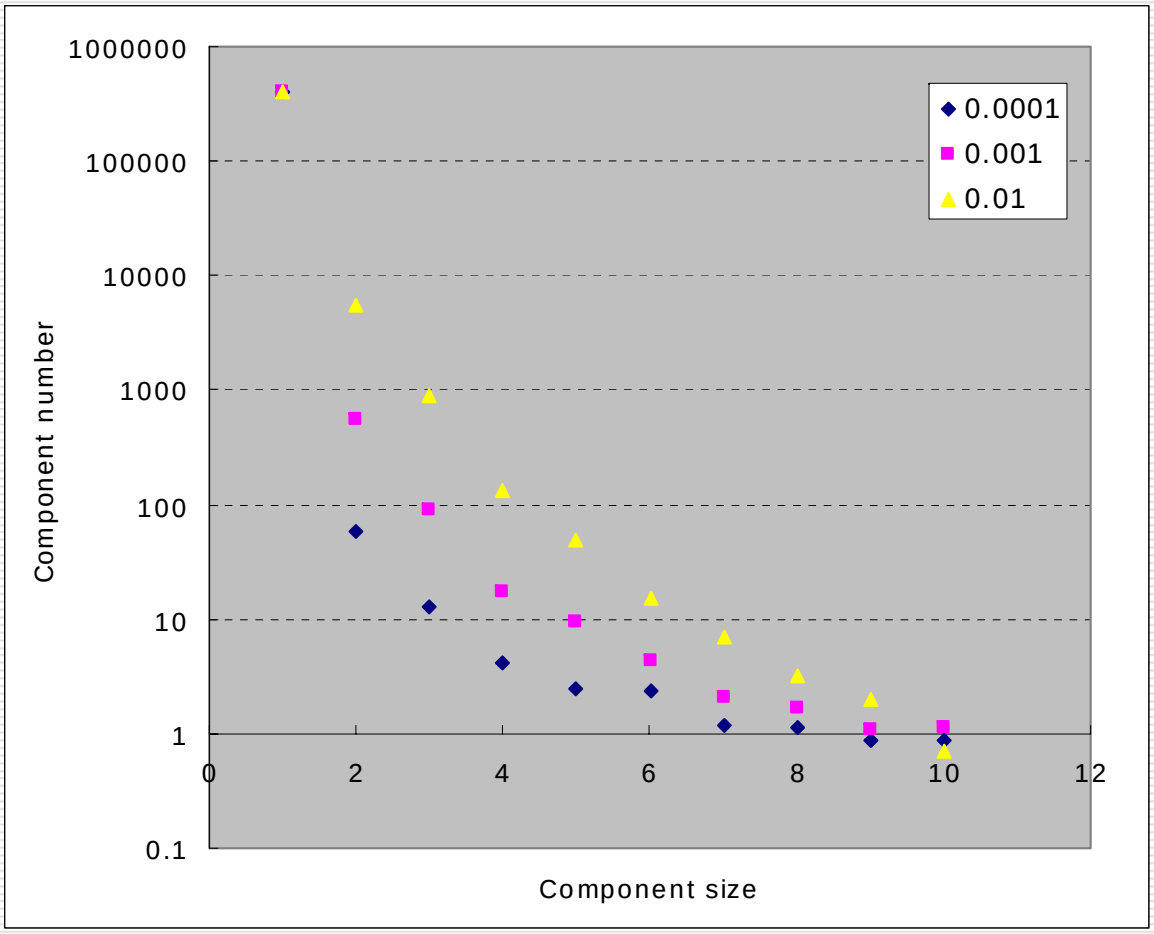


Average degree <k> v.s. step for evolving networks with different mutation rates (P_{mut} = 0.01, 0.001, 0.0001)

* Data shown are means over 10 graphs, from 100 (N_0) nodes to about 400,000 (N_{13}) nodes

A step means the whole progress of duplication and divergence in GPC model.

GPC model: Component size distribution



Size distribution of connected components in a network, evolving with different mutation rates (P_{mut} = 0.01, 0.001, 0.0001)

* Data shown are means over 10 graphs of about 400,000 (N_{13}) nodes

The size means the number of vertices in each connected component.

Discussion

- The mutation rate influences the topology of gene-protein contact networks
  : Considering the real time scale (e.g. using the synonymous substitution rate, K_s)

- Comparison with random graphs
  : Hard to find suitable random graphs

- High performance computation needed
  : Examination of topology of growing networks
  (Example) Simulating growing graphs with 2 million nodes within 1 min.

Development environment

- C++ computer language
  : Most suitable for high-performance computation

- The C++ Boost Graph Library
  : http://www.boost.org/
  : Using generic programming to provide general purpose graph classes and interfaces

- Graph visualization software
  Graphviz - http://www.graphviz.org/
  ZGRViewer - http://zvtm.sourceforge.net/

References

Albert, R. and A.L. Barabasi. 2000. Statistical mechanics of complex networks. *Rev Mod Phys* 85: 5234-5237.

Barabasi, A.L. and Z.N. Oltvai. 2004. Network biology: Understanding the cell's functional organization. *Nat Rev Genet* 5: 101-U115.

Duarte, J.M., L. Cui, P.K. Wall, Q. Zhang, X. Zhang, J. Leebens-Mack, H. Ma, N. Altman, and C.W. dePamphilis. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. *Mol Biol Evol* 23: 469-478.

Farid, N. and K. Christensen. 2006. Evolving networks through deletion and duplication. *New J Phys* 8: 1-18.

Force, A., M. Lynch, F.B. Pickett, A. Amores, Y.L. Yan, and J. Postlethwait. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531-1545.

Ohno, S. 1970. Evolution by gene duplication. Springer,NewYork.

Pastor-Satorras, R., E. Smith, and R.V. Sole. 2003. Evolving protein interaction networks through gene duplication. *J Theor Biol* 222: 199-210.

Wagner, A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol* 18: 1283-1292.

Wendel, J.F. 2000. Genome evolution in polyploids. *Plant Mol Biol* 42: 225-249.